

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:40:53 ; Search time 3416.8 Seconds
(without alignments)
7998.346 Million cell updates/sec

Title: US-09-896-888A-1

Perfect score: 564

Sequence: 1 catgatgataacaatgtat.....tggtagcgacacaaacatg 564

Scoring table: IDENTIFY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	564	100.0	564	6	BD070856		BD070856 Insect ex
2	564	100.0	4170	14	NPHTTAA		M83827 Orgyia pseu
3	564	100.0	131995	14	OPU75930		U75930 Orgyia pseu
C 4	562.4	99.7	1429	14	S64501		S64501 p8.9=8.9 kd
5	548	97.2	2773	6	AX766573		AX766573 Sequence
C 6	111.8	19.8	118584	14	AY043265		AY043265 Epiphyas
C 7	101.8	18.0	131526	14	A1145471		A1145471 Rachiphus
C 8	101.4	18.0	1511	14	NPHPE38		M62488 Autographa
C 9	101.4	18.0	133894	6	A48542		A48542 Sequence 1
C 10	101.4	18.0	133894	14	L22858		L22858 Autographa
C 11	100	17.7	131158	14	AY327402		AY327402 Choriston
C 12	93.4	16.6	28413	6	BD187790		BD187790 A virus i
C 13	93.4	16.6	128413	14	NPHTT3COMP		L33180 Bombyx mori
C 14	92.2	16.3	129609	14	AF512031		AF512031 Choriston
C 15	92	16.3	2178	14	NPBEMTEN		D14467 Bombyx mori
C 16	90.4	16.0	2011	14	NPHIEN		M59422 Autographa
17	73.8	13.1	155060	14	MBU59461		U59461 Mamestra co
18	72.2	12.8	4051	14	AF246708		AF246708 Spodopte
C 19	72.2	12.8	139342	14	AF325155		AF325155 Spodopte

20	72.2	12.8	153656	14	AF539999	AF539999 Mamestra
C 21	69.2	12.3	15528	14	AF107100	AF107100 Ecotropla
22	68.2	12.1	158482	14	AY126275	AY126275 Mamestra
C 23	67.4	12.0	130759	14	AF303045	AF303045 Helicover
C 24	66.4	11.8	14235	14	HZU67264	U67264 Helicoverpa
C 25	65.6	11.6	130869	14	AF334030	AF334030 Helicover
C 26	65.4	11.6	135611	14	AF169823	AF169823 Spodopte
C 27	64	11.3	131403	14	AF271059	AF271059 Helicove
C 28	63.6	11.3	161046	14	AF081810	AF081810 Lymantria
C 29	59.2	10.5	113220	14	AP006270	AP006270 Adoxophye
C 30	41.8	7.4	131158	14	AY327402	AY327402 Choriston
C 31	40.2	7.1	128380	2	AC151093	AC151093 Bos tauru
C 32	39.6	7.0	6289	6	AX598859	AX598859 Sequence
C 33	39.6	7.0	9289	6	AX251255	AX251255 Sequence
C 34	39.6	7.0	9289	6	AX767469	AX767469 Sequence
C 35	39	6.9	152573	2	CR376837	CR376837 Danio rer
C 36	39	6.9	200989	5	BX000999	BX000999 Zebrafish
C 37	38	6.7	6289	6	AX599005	AX599005 Sequence
C 38	38	6.7	9289	6	AX767545	AX767545 Sequence
C 39	38	6.7	160376	4	AC150919	AC150919 Bos tauru
C 40	37.4	6.6	96824	2	AC020384	AC020384 Drosophil
41	36.6	6.5	125422	9	AL449063	AL449063 Human DNA
C 42	36.6	6.5	175909	9	AC093559	AC093559 Homo sapi
C 43	36.6	6.5	190808	10	AL928537	AL928537 Mouse DNA
C 44	36.6	6.5	221128	10	AC123846	AC123846 Mus muscu
C 45	36.4	6.5	505	6	AR424142	AR424142 Sequence

ALIGNMENTS

RESULT 1
BD070856
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BD070856
Insect expression vectors.
BD070856
BD070856.1 GI:22616459
JP 2001516225-A/1.
unidentified
unidentified
unclassified.
1 (bases 1 to 564)
Grigliatti,T.A., Theilmann,D.A., Pfeifer,T.A. and Hegedus,D.D.
Insect expression vectors
Patent: JP 2001516225-A 1 25-SEP-2001;
THE UNIVERSITY OF BRITISH COLUMBIA
OS Multicapsid nucleopolyhedrovirus
PN JP 2001516225-A/1
PD 25-SEP-2001
PF 26-MAR-1998 JP 1998541010
PR 27-MAR-1997 US 60/049946,28-JAN-1998 CA 2221819 PI
THOMAS A GRIGLIATTI,DAVE A THEILMANN,THOMAS
A PFEIFER,DWAYNE D
PI HEGEDUS
PC C12N15/06,C12N15/69//C12N9/22
CC Insect expression vectors
FH Key Location/Qualifiers
FT source 1..564
FT /organism='Multicapsid nucleopolyhedrovirus'.
Location/Qualifiers
1..564
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'

FEATURES

source

ORIGIN

Query Match 100.0%; Score 564; DB 6; Length 564;
Best Local Similarity 100.0%; Pred.No.4e-171;
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGATGATAACATGTATGTGCTTAATGTTGCTTCAACACAATTCGTGTAACCTG 60
DB 1 CATGATGATAACATGTATGTGCTTAATGTTGCTTCAACACAATTCGTGTAACCTG 60

QY 61 TTTTCATGTTTGGCCAAACAGCACCTTTTATATCTCGGTGGCTCCCAACCACTTTT 120
Db 61 TTTTCATGTTTGGCCAAACAGCACCTTTTATATCTCGGTGGCTCCCAACCACTTTT 120
QY 121 GCACTGCAAAAAAACAACACCTTTTGGCAGCGGGCCCATACATAGTACAACTCTACGTTTC 180
Db 121 GCACTGCAAAAAAACAACACCTTTTGGCAGCGGGCCCATACATAGTACAACTCTACGTTTC 180
QY 181 GTAGACTATTTTACATAAATAGTCTACACCTTGTTATACGTTGATACGTTCCAAATACACTACCAAC 240
Db 181 GTAGACTATTTTACATAAATAGTCTACACCTTGTTATACGTTGATACGTTCCAAATACACTACCAAC 240
QY 241 ATTGAACCTTTTTCAGTGCAGGCAAAAAGTACGTTGCGGAGTACGTTAGCGGCGCTTATC 300
Db 241 ATTGAACCTTTTTCAGTGCAGGCAAAAAGTACGTTGCGGAGTACGTTAGCGGCGCTTATC 300
QY 301 GGGTCCGCTCTGTCACGTCAGCAATACATATCGGACCGGACGAGTGTCTTATCGT 360
Db 301 GGGTCCGCTCTGTCACGTCAGCAATACATATCGGACCGGACGAGTGTCTTATCGT 360
QY 361 GACAGGACCGCAGCTTCTGTGTTGCTAAACCGGACCGGACGCACTCTTATCGGAACA 420
Db 361 GACAGGACCGCAGCTTCTGTGTTGCTAAACCGGACCGGACGCACTCTTATCGGAACA 420
QY 421 GGACGCGCTCCATATCAGCGCGGTTTATCTCATGCGGTCAGCGGACGAGCGGCC 480
Db 421 GGACGCGCTCCATATCAGCGCGGTTTATCTCATGCGGTCAGCGGACGAGCGGCC 480
QY 481 GTCCCGCTTATCGGCGCTTATAAATACAGCCCGCAACGATCTGGTAAACACAGTGAACAG 540
Db 481 GTCCCGCTTATCGGCGCTTATAAATACAGCCCGCAACGATCTGGTAAACACAGTGAACAG 540
QY 541 CATCTGTTACAGCGACACCAACATG 564
Db 541 CATCTGTTACAGCGACACCAACATG 564

RESULT 2
NPHTTAA 4170 bp DNA linear VRL 02-AUG-1993
LOCUS
DEFINITION
Orgyia pseudotsugata nuclear polyhedrosis virus transcriptional
trans-activator (IE-2) gene, complete cds; ORF, complete cds.
ACCESSION
M83827
VERSION
M83827.1 GI:332540
KEYWORDS
transcriptional transactivator.
SOURCE
Orgyia pseudotsugata single capsid nucleopolyhedrovirus
Orgyia pseudotsugata single capsid nucleopolyhedrovirus
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE
1 (bases 1 to 4170).
AUTHORS
Theilmann,D.A. and Stewart,S.
TITLE
Molecular analysis of the trans-activating IE-2 gene of Orgyia
pseudotsugata multicapsid nuclear polyhedrosis virus
JOURNAL
Virology 187 (1), 84-96 (1992)
MEDLINE
92142536
PUBMED
1736546
REFERENCE
2 (bases 1 to 4170)
AUTHORS
Theilmann,D.A. and Stewart,S.
TITLE
Randomly repeated sequence at the 3' end of the IE-2 gene of the
baculovirus Orgyia pseudotsugata multicapsid nuclear polyhedrosis
virus is an enhancer element
JOURNAL
Virology 187 (1), 97-106 (1992)
MEDLINE
92142537
PUBMED
1736547
COMMENT
Original source text: Orgyia pseudotsugata nuclear polyhedrosis
virus DNA.
FEATURES
Location/Qualifiers
source
1..4170
/organism="Orgyia pseudotsugata single capsid
nucleopolyhedrovirus"
/mol_type="genomic DNA"
/db_xref="taxon:10450"

gene 627..3959
TATA_signal
/gene="IE-2"
627..634
/gene="IE-2"
mRNA
657..1908
/gene="IE-2"
691..1908
/gene="IE-2"
CDS
/codon_start=1
/product="transcriptional trans-activator"
/protein_id="AAA46749.1"
/db_xref="GI:332541"
/translation="MSPSNNAAPTSPNRRNLSLVGRRLTYSPPDAASAQASPPR
SAPRAAPRVHAGVDCGAPLRASVALPNGVYNLHGDAHFPPEDDDILFVDTAAEQ
RQAVNUHEAVNRHURRGLGERMTRSPILLNYSPTSPSRSPDLIMPDLQ
ARQPLAFNDSDDDERLEQVMSAEAPQLQAPQAPQVDVSLVCHICSCITDIO
NYSNFTSTECNHAVCKYVSIVFKGSKYKSCINRRTISCRAYNRDGVSLTMS
TVNDSQAIKRHWQLSDSNPHSNEMTTIOELQALAEALRAATARAHDVNMARSDS
LLRQLOVKEABELAHESNARLKQKQNETLSANNLSLQHLQNTQVIESRVKQEPKQ
HDEPMEXFKLSLS"
polyA_signal
1958..1963
/gene="IE-2"
enhancer
2017..2759
/gene="IE-2"
TATA_signal
2915..2922
/gene="IE-2"
mRNA
2944..3907
/gene="IE-2"
CDS
2984..3907
/gene="IE-2"
/note="ORF: homologous to the ACMPV PE-38 gene; putative"
/codon_start=1
/protein_id="AAA46750.1"
/db_xref="GI:332542"
/translation="MSSRYSPPYRIRNVGERRRVOERLLSEFNSRPVTAVAACVCL
TYCVQSNIIIDFLMPSCTHLCFCVKVNNYKNMNVPRAAVSCPMCKNKVGTQWQAF
PNSVSKKIKTGDRTDPAQLFALKTQDRYMAEEAEPTSFVKNLOALDA
AOKEARDLOESMERQKOHNVNNSCEQVTLQTLADQALDRSEALSLAEHN
RAANVQIDSLRRVORLEAAQSPVSVNVFENDNQNTNLHERFRYVYSTVSDMI
EDTSKISQSHVFGAACLPCSVNVEINPFDE"
polyA_signal
3954..3959
/gene="IE-2"
ORIGIN
Query Match 100.0%; Score 564; DB 14; Length 4170;
Best Local Similarity 100.0%; Pred. No. 6.1e-171;
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGATGATAAACAATGTATGGTCTAATGTGCTTCAACAACAATCTGTTGAACCTGTG 60
Db 130 CATGATGATAAACAATGTATGGTCTAATGTGCTTCAACAACAATCTGTTGAACCTGTG 189
QY 61 TTTTCATGTTTGGCAACAAGCACCTTTTATATCTCGGTGGCTCCCAACCACTTTT 120
Db 190 TTTTCATGTTTGGCAACAAGCACCTTTTATATCTCGGTGGCTCCCAACCACTTTT 249
QY 121 GCACGTCAAAAAAACAACGCTTTTGCAGCGGGCCCATACATAGTACAACTCTACGTTTC 180
Db 250 GCACGTCAAAAAAACAACGCTTTTGCAGCGGGCCCATACATAGTACAACTCTACGTTTC 309
QY 181 GTAGACTATTTTACATAAATAGTCTACACCTTGTTATACGTTCCAAATACACTACCAAC 240
Db 310 GTAGACTATTTTACATAAATAGTCTACACCTTGTTATACGTTCCAAATACACTACCAAC 369
QY 241 ATTGAACCTTTTTCAGTGCAGGCAAAAAGTACGTTGCGGAGTACGTTAGCGGCGCTTATC 300
Db 370 ATTGAACCTTTTTCAGTGCAGGCAAAAAGTACGTTGCGGAGTACGTTAGCGGCGCTTATC 429
QY 301 GGGTCCGCTCTGTCACGTCAGCAATACATATCGGACCGGACGAGTGTCTTATCGT 360
Db 430 GGGTCCGCTCTGTCACGTCAGCAATACATATCGGACCGGACGAGTGTCTTATCGT 489
QY 361 GACAGGACCGCAGCTTCTGTGTTGCTAAACCGGACCGGACGCACTCTTATCGGAACA 420

```
Db 490 GACAGGCGCCAGCTTCTGTGTTCTAACCAGCGCGCAACTCTTATCGGAACA 549
Qy 421 GGACCGCGCTCCATATGACCGCGGTTTATCTCATGCGGTGACCGGACACGAGCGGCC 480
Db 550 GGAAGCGCTCCATATGACCGCGGTTTATCTCATGCGGTGACCGGACACGAGCGGCC 609
Qy 481 GTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAACACAGTTGAACAG 540
Db 610 GTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAACACAGTTGAACAG 669
Qy 541 CATCTGTTACAGCGACACAAATG 564
Db 670 CATCTGTTACAGCGACACAAATG 693

RESULT 3
OPU75930 131995 bp DNA circular VRL 25-MAR-2003
LOCUS Orygia pseudotsugata multicapsid nucleopolyhedrovirus, complete
DEFINITION
ACCESSION U75930
VERSION U75930.2 GI:11024985
KEYWORDS
SOURCE Orygia pseudotsugata multicapsid nucleopolyhedrovirus
ORGANISM Orygia pseudotsugata multicapsid nucleopolyhedrovirus
          Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
          Nucleopolyhedrovirus.
REFERENCE 1 (bases 1 to 131995)
AUTHORS Ahrens,C.H., Russell,R.L., Funk,C.J., Evans,J.T., Harwood,S.H. and
          Rohrmann,G.F.
TITLE The sequence of the Orygia pseudotsugata multinucleocapsid nuclear
JOURNAL polyhedrosis virus genome
MEDLINE Virology 229 (2), 381-399 (1997)
PUBMED 97271300
REFERENCE 2 (bases 1 to 131995)
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1996) Oregon State University, Agricultural
          Chemistry, Corvallis, OR 97331-7301, USA
REFERENCE 3 (bases 1 to 131995)
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1998) Oregon State University, Agricultural
          Chemistry, Corvallis, OR 97331-7301, USA
REMARK Sequence update by submitter
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2000) Oregon State University, Agricultural
          Chemistry, Corvallis, OR 97331-7301, USA
REMARK Sequence update by submitter
COMMENT On Oct 26, 2000 this sequence version replaced gi:2934903.
FEATURES
          location/Qualifiers
            1..131995
              /organism="Orygia pseudotsugata multicapsid
              nucleopolyhedrovirus"
              /mol_type="genomic DNA"
              /db_xref="taxon:164623"
              /note="similar to Autographa californica nuclear
              polyhedrosis virus (AcMNPV) deposited in GenBank Accession
              Number L22858"
            complement(123..947)
              /note="ORF1; pk-1; similar to AcMNPV ORF10"
            /codon_start=1
              /product="protein kinase 1"
              /protein_id="AAC59000.1"
              /db_xref="GI:1911247"
              /translation="MDAALQSLRSFVADCAVLPAKVNGRFGKMDVLRHPTTSKLF
              RKTAAHFSADEINVDLMDSHPSFDMYFCYSSTPAWIMDVVPCPDLPETLQTO
              GAIADNALVNIUROLCDALNDLNATGYIHNDVKLENVLYFGARDRVLYCDGLCKRE
              HSPVHDGTELYFSPKEIRRHNYARSFDWYAVGVLAUKLTGGRHFFERSADEVDLAS
```

CDS

```
MRRQQVNDPAALKNVNLMMARDVFVCLTRFNFCRSTDYKQIAKHSFLASRHDYI"
946..2370
/note="ORF2; similar to AcMNPV ORF9"
/codon_start=1
/product="1629-capsid"
/protein_id="AAC59001.1"
/db_xref="GI:2934904"
/translation="MMERQYOSVRSYLYINNOHNAIAAGPFLQVAGPBAHSGVGRNVGD
RVLNRAQVLDLLKLAEDIYADTA YMQADQPEASSSRHFATLNRMLLLIGVQDPAR
RNLNSVLRIEALLRVVDVNDVDAENVLSGDFEYSKYISYQTAQTPTASASQQTQ
TSLPRPOTSLPQOTPFQDQPMVSPSPVHTTTPAI LPOTTQPPATDTPSRP
SDEFVYVGKRAVEDTRFKPPKPEHLKASRSPSSVATNAAGATFVAPPPPSAD
VTTSMPPPPPPPSADVTTSMPPPPPPVVDLATSMPPPPPPVVDLATSMPPPINNA
INNLIDAMVAETNKNAGDNRSLDQIKQKGLTKKTQADCAPATDPRSTLLSEIR
GKTLKLRKI BDQSTQTLLKDVDTTKTKILKNFVTDNRISKQOEKEDRLDTIT
KRRPAVEHTDGNSTGNNSDDWRDD"
complement(2533..3270)
/note="ORF3; ph; similar to AcMNPV ORF8"
/codon_start=1
/product="polyhedrin"
/protein_id="AAC59002.1"
/db_xref="GI:1911249"
/translation="MPDYSYRPTIGRTVYDNKYKNLGSVIKNAKKKHLEHEEDE
KHLDPLOYMVAEDPFLGPGKNOKLTLFKEIRNVKPDTKLIVNWSGFEURETFRF
VEDSFPVNDQEVNDVFLVNNRTPRNCYKFLAQHALRWDCCDVPHVIRIVPSY
VGMNNEYRISLAKGGGCPINNIHAEYTNSEFSEFVNVIWENFYKPIVIYIGDSSEE
BILIEVSLVFKVGFAPDAPLFTGPAY"
complement(3350..3541)
/note="ORF4"
/codon_start=1
/product="unknown"
/protein_id="AAC59003.1"
/db_xref="GI:1911250"
/translation="MRKNGASQYPEHLTTVHKRRLQLSCVWCDDQDSDWDPHSAGKL
MEHMFCLRIYVOQIKH"
complement(3698..4093)
/note="ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAC59004.1"
/db_xref="GI:1911251"
/translation="MGASONSEOKILKLFYWSQTAALDDEKDLHCLYDLERFVGA
HLNKRADDKARKKCAEAKALKRVEIAADRHMLEAANAACADDGRWSTLSRAQLDDI
AREKEIVDRHRLQLKQDSLLKQDRLLKR"
complement(4125..4739)
/note="ORF6; left-2; similar to AcMNPV ORF6"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAC59005.1"
/db_xref="GI:1911252"
/translation="MERVWNPAAIGDLKRSETYLVDPHFVGLTSPYTVFERGLF
VRMSGMLLALLAAPKQEPQVAPRRFPQSRNRNVLKACADQAQSLAKVLAARYSMP
PCMSKTWADLSSAPGNMYRRFENCYLANVITCTKCTACILGALLHFRYMDAKCV
GEVTHLLIKAQDVYKPSNCAEMKKVKLCPQASMKGLNPICNF"
complement(4742..4975)
/note="ORF7; similar to AcMNPV ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAC59006.1"
/db_xref="GI:1911253"
/translation="MNRFTMENTAAVTTDYDREQLRRLNSLRSRVHSLCTRSTATGFD
CNRFLKADRAPAVIVKAAANGGQHSLLICDKV"
5020..5460
/note="ORF8; similar to AcMNPV ORF4"
/codon_start=1
/product="unknown"
/protein_id="AAC59007.1"
/db_xref="GI:1911254"
/translation="MPGARFVRFLRLTQEFKENVAVHDLMLGRALIDGKVTSAV
RRFGFLSNVLSVSAAMVNVQVYAPDATIDMRHQPTIYFRVCQNCAMADVPAPDHS
IARYLLACGAVLVIDHPLDVFGETEGWNLLEQVINAGDGL"
complement(5669..6151)
/note="ORF9; ptp-2"
```

CDS

CDS

```
/codon_start=1
/product="protein tyrosine phosphatase 2"
/protein_id="AAC59008.1"
/db_xref="GI:1911255"
/translacion="MYEASQIDHFVGGYGGYDNEAMLFENHAIASVISLSDVA
PIRHALGLPVGDIHVCCEAAPTCAALPNAMPALYDVMVRRIIGSGKRVLVHCYAGASR
SAALAVVYLMESRQMPYQDALNAVQSKRRVAINDHFVFLATRCYSYFVNDLKPQIV
"
```

```
complement(6129..6791)
/note="ORF10; ptp-1; similar to AcMNPV ORF1"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAC59009.1"
/db_xref="GI:1911256"
/translacion="MLTSRWAVIYTRTSRNCGGAWCTGVRRRRQVHLPSVFRAM
FPDRHBYTCAGVITRLLCFKYNLAELFEYVTSDEDRWTAASVLARHSALGAVI
DLNTARYDGAQVKNGLYKKIRVFGRAVPDDDIIVAEFIETVDEFRRCPMLNAV
HWTHLNRSGYLVCYVVERLGVSPDTDAIARFETARGHKIERTNYLQDLLARKHVRQ
PN"
```

```
6819..7814
/note="ORF11; similar to AcMNPV ORF11"
/codon_start=1
/product="unknown"
/protein_id="AAC59010.1"
/db_xref="GI:1911257"
/translacion="MSLSKLLVYVGSYNLPHDRYGESYHLVRIIVHEHLTYVSN
ASVRRDIATACLSNHLCPDVARQLLDVSEVNAARLSAIFRPGDGTGLCADMORALA
DIDHAPLARVRNIFALDAIDIPSDVNNLQGIIGRFMHFPRCSGLARVADVF
DIPRAGDWMYKFCVLTMYHLVACGAPAGSATRLRDVAKHIGPNDEGNCAPIAA
VYGRCAIGREHFAHTACMHLIFQFMRNDLTTPADERHPCFVIGDKFGRQCKDTYD
LRTHADALYHGTTRDOKNALFDLLCCVNASDIDADCYCVNKFYATQNKYKM"
complement(8180..9142)
```

```
/note="ORF12; similar to AcMNPV ORF13"
/codon_start=1
/product="unknown"
/protein_id="AAC59011.1"
/db_xref="GI:1911258"
/translacion="MFESMFGWNAADQVNAEFDQAYRYAVDQRAHSDLVRRDVP
RCHFVFPRYVLDIATAGRCSSVDFCKGLKISHDLLQRCNDFRQHVRLNVLGNP
PAQDSIGSUFATKHGLVQLLQFSPFANKNEVLLAANGKNDHDLNLDKIEAVLHAP
KTLNNSDKFTISAHKSPKFEVGFQRLQTLTKLALQCAAPTRTAPGVVFPFR
DVTKPHLAVFMVDEVRGVTQIAFARGQBEHFRKLEFEEGMDVDVGRAPNPLIA
VHCIEEFGANGHKIRLPPKVIIEVDCAVNAKDIVKKAINKT"
complement(9094..9825)
```

```
/note="ORF13; left-1; similar to AcMNPV ORF14"
/codon_start=1
/product="late expression factor 1"
/protein_id="AAC59012.1"
/db_xref="GI:1911259"
/translacion="WAPCKYTPERVOMMDAIAVNSRRLAEMTDRPRWVHAHNFDS
AAQLFAYIKNSISDVHVKLEEGREWVIDADPKDCADKAELMLKYNVGTATPMLFP
EGKEDAVQRMFSGNRFHMLKFCGFKMDAPKSLRHNWPNFKPAKLVGSDIRPG
SFQDCVRAVHMYIGDAREDLVLRVDPDRDVEFCNANKQIRAPFSYNGGDSYRCL
TQQLQRIKACACAGAGTPTTSK"
9893..11362
```

```
/note="ORF14; egt; similar to AcMNPV ORF15"
/codon_start=1
/product="ecdysteroid UDP-glucosyltransferase"
/protein_id="AAC59013.1"
/db_xref="GI:1911260"
/translacion="MVFILALTLATGARAASLAVLPTPAYSHVYVYAYVHALVK
NCHNVTVIKOLLVAVQDEGRVEQIDADMSAQYKKLVASSGVRKRGVADETTV
TADNYMGLIEMKQDFQNAVVRFLSTNRTPDAVVEAPADYALVFGHLPAPVIOI
APYGLAENERRKARVAPRHPLFTFGAALTRRGGLSEWRLNLFELLARSDGLL
KQOKESTPTIRQVNDVQGLLNLHPVNNRVPSPVQVYLGGLHLAQLPORDA
PFLERNSVDQVAVFGSIDTNSHAEPQLMLDTFANLNNYTLVWYKVVDAVAS
VALPNLAQWFSQVAVLRKNVAVFTQAGLSDSEALQAPVPMVCLPMWGQDFH
ARKLQOFGVARALTAASVAPOLQAIREVIADEGEAYRATDKLRVDEHDAAPDEKA
VKFTERTVKNNDVNWPARSLKTTAANMAYSDYVRFPL"
11547..12140
```

```
/note="ORF15; similar to AcMNPV ORF16"
/codon_start=1
```

```
Query Match 100.0%; Score 564; DB 14; Length 131995;
Best Local Similarity 100.0%; Pred. No. 1.3e-170;
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CATGATCATAAACAATGATAGGTGCTTAATGTCTTCAACAACAATTTCTGTGAACCTGTG 60
Db 128172 CATGATCATAAACAATGATAGGTGCTTAATGTCTTCAACAACAATTTCTGTGAACCTGTG 128231
Qy 61 TTTTCATGTTTGGCAACAAGCACTTTTATCTGGTGGCTCCCAACCACTTTT 120
Db 128232 TTTTCATGTTTGGCAACAAGCACTTTTATCTGGTGGCTCCCAACCACTTTT 128291
Qy 121 GCATGCAAAAAAACACGCTTTTTCACGCGGGCCATACATAGTACAAACTCTACGTTTC 180
Db 128292 GCATGCAAAAAAACACGCTTTTTCACGCGGGCCATACATAGTACAAACTCTACGTTTC 128351
Qy 181 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATACCGTCCAAATACACTACCAC 240
Db 128352 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATACCGTCCAAATACACTACCAC 128411
Qy 241 ATTGAACCTTTTTCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 300
Db 128412 ATTGAACCTTTTTCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 128471
Qy 301 GGGTCGGCTCTGTCAGTACGAATCACTTATCGGACCGGACGAGTGTGTCTTATCGT 360
Db 128472 GGGTCGGCTCTGTCAGTACGAATCACTTATCGGACCGGACGAGTGTGTCTTATCGT 128531
Qy 361 GACAGGACGCGAGCTTCTGTGTTGCTAACCGGACCGGACGCAACTCCTTATCGGAACA 420
Db 128532 GACAGGACGCGAGCTTCTGTGTTGCTAACCGGACCGGACGCAACTCCTTATCGGAACA 128591
Qy 421 GGACGCGCTTCATATCAGCGCGCGCTTATCTATCGGTCGACGACGACGACGCGCGCC 480
Db 128592 GGACGCGCTTCATATCAGCGCGCGCTTATCTATCGGTCGACGACGACGACGCGCGCC 128651
Qy 481 GTCCCGCTTATCGCGCTTAAATACAGCGCGCAACATCTGGTAAACACAGTTGAACAG 540
Db 128652 GTCCCGCTTATCGCGCTTAAATACAGCGCGCAACATCTGGTAAACACAGTTGAACAG 128711
Qy 541 CATCTGTTACAGGACACACATG 564
Db 128712 CATCTGTTACAGGACACACATG 128735
```

RESULT 4

S64501/c

LOCUS

DEFINITION

S64501

nuclear polyhedrosis virus OpMNPV, Genomic, 1429 nt).

ACCESSION

S64501

VERSION

S64501.1

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REMARK

FEATURES

source

1..1429

/organism="Orgyia pseudotsugata multicapsid

nucleopolyhedrovirus"

/mol_type="genomic DNA"

/db_xref="taxon:164623"

S64501 1429 bp DNA linear VRL 30-SEP-1993
p8.9=8.9 kda basic protein [Orgyia pseudotsugata multicapsid
nuclear polyhedrosis virus OpMNPV, Genomic, 1429 nt].

S64501 GI:404518
S64501.1

Orgyia pseudotsugata multicapsid nucleopolyhedrovirus
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.

1 (bases 1 to 1429).
Wu, X., Stewart, S. and Theilmann, D.A.
Characterization of an early gene coding for a highly basic 8.9K
protein from the Orgyia pseudotsugata multicapsid nuclear
polyhedrosis virus

J. Gen. Virol. 74 (Pt 8), 1591-1598 (1993)

93346965
8345350
GenBank staff at the National Library of Medicine created this
entry [NCBI gibseq 136152] from the original journal article.

Location/Qualifiers
1..1429

/organism="Orgyia pseudotsugata multicapsid
nucleopolyhedrovirus"

/mol_type="genomic DNA"
/db_xref="taxon:164623"


```

gene      763..990
/ gene="p8.9"
/ note="8.9 kda basic protein"
CDS       763..990
/ gene="p8.9"
/ note="8.9 kda basic protein"
/ codon_start=1
/ protein_id="AAB27738.1"
/ db_xref="GI:404519"
/ translation="MNSQKIRMAKQQQVRVARQHRRAKGLRLYKAKKLRAELCEKLG
LQRVNDAALAKAPEEFVPHFSFLYTLN"

ORIGIN
Query Match      99.7%; Score 562.4; DB 14; Length 1429;
Best Local Similarity 99.8%; Pred. No. 1.6e-170;
Matches 563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CATGATGATAAACAATGATGCTGCTAATGTTGCTTCAACACAATTCGTGTAACCTGTG 60
Db 765 CATGATGATAAACAATGATGCTGCTAATGTTGCTTCAACACAATTCGTGTAACCTGTG 706

QY 61 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 120
Db 705 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 646

QY 121 GCACGCAAAAAAACAACGCTTTTGCACGGGGCCCATACATAGTACAACTCTAGCTTTC 180
Db 645 GCACGCAAAAAAACAACGCTTTTGCACGGGGCCCATACATAGTACAACTCTAGCTTTC 586

QY 181 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTCCCAATACACTACACAC 240
Db 585 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTCCCAATACACTACACAC 526

QY 241 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGGAGTCACGTAGCGCGGCTTATC 300
Db 525 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGGAGTCACGTAGCGCGGCTTATC 466

QY 301 GGGTCGGCTCTGTACGATGCAAAATCAATATCGGACCGGACGAGTGTGCTTATCGT 360
Db 465 GGGTCGGCTCTGTACGATGCAAAATCAATATCGGACCGGACGAGTGTGCTTATCGT 406

QY 361 GACAGGACCGCAGCTTCTGTGCTTCAACCGGACGAGTGTGCTTATCGGACCA 420
Db 405 GACAGGACCGCAGCTTCTGTGCTTCAACCGGACGAGTGTGCTTATCGGACCA 346

QY 421 GACAGCGGCTTCCATATCAGCGCGGCTTATCTCATGCGGTCACCGGACGAGCGGCC 480
Db 345 GACAGCGGCTTCCATATCAGCGCGGCTTATCTCATGCGGTCACCGGACGAGCGGCC 286

QY 481 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTGAACAG 540
Db 285 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTGAACAG 226

QY 541 CATCTGTTACAGCGACACAACTG 564
Db 225 CATCTGTTACAGCGACACAACTG 202

RESULT 5
LOCUS      AX766573
DEFINITION Sequence 60 from Patent WO03042244.
ACCESSION  AX766573
VERSION     AX766573.1 GI:32260450
KEYWORDS   .
SOURCE      synthetic construct
            other sequences; artificial sequences.
ORGANISM    1
REFERENCE   1
AUTHORS     Klysner,S., Nielsen,F.S., Bratt,T., Voldborg,B. and Mouritsen,S.
TITLE       Novel immunogenic mimetics of multimer proteins
JOURNAL     Patent: WO 03042244-A 60 22-May-2003;
            Pharmexa A/S (DK) ; Klysner, Steen (DK) ; Nielsen, Finn Stausholm

```

```

(DK) ; Bratt, Tomas (DK) ; Voldborg, Bjorn (DK) ; Mouritsen, Soren
(DK)

FEATURES             source
Location/Qualifiers
1..2773
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/ note="p2Op2F expression vector for insect cells"
561..566
/ note="HindIII site"
573..578
/ note="Aval site"
586..591
/ note="EcoRI site"
593..598
/ note="BamHI site"
625..630
/ note="ClaI site"
629..634
/ note="ClaI site"
1156..1161
/ note="ApaLI site"
2128..2133
/ note="PstI site"
2204..2209
/ note="NcoI site"
2284..2289
/ note="Aval site"
2294..2299
/ note="Aval, SmaI, and XmaI site"
2551..2556
/ note="ApaLI site"

ORIGIN
Query Match      97.2%; Score 548; DB 6; Length 2773;
Best Local Similarity 100.0%; Pred. No. 8.6e-166;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGATGATAAACAATGATGCTGCTAATGTTGCTTCAACACAATTCGTGTAACCTGTG 60
Db 5 CATGATGATAAACAATGATGCTGCTAATGTTGCTTCAACACAATTCGTGTAACCTGTG 64

QY 61 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 120
Db 65 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 124

QY 121 GCACGCAAAAAAACAACGCTTTTGCACGGGGCCCATACATAGTACAACTCTAGCTTTC 180
Db 125 GCACGCAAAAAAACAACGCTTTTGCACGGGGCCCATACATAGTACAACTCTAGCTTTC 184

QY 181 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTCCCAATACACTACACAC 240
Db 185 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTCCCAATACACTACACAC 244

QY 241 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGGAGTCACGTAGCGCGGCTTATC 300
Db 245 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGGAGTCACGTAGCGCGGCTTATC 304

QY 301 GGGTCGGCTCTGTACGATGCAAAATCAATATCGGACCGGACGAGTGTGCTTATCGT 360
Db 305 GGGTCGGCTCTGTACGATGCAAAATCAATATCGGACCGGACGAGTGTGCTTATCGT 364

QY 361 GACAGGACCGCAGCTTCTGTGCTTCAACCGGACGAGGCAACTCTCTTATCGGAACA 420
Db 365 GACAGGACCGCAGCTTCTGTGCTTCAACCGGACGAGGCAACTCTCTTATCGGAACA 424

QY 421 GGACGGGCTTCATATCAGCGCGGCTTATCTCATGCGGTCACCGGACGAGCGGCC 480
Db 425 GGACGGGCTTCATATCAGCGCGGCTTATCTCATGCGGTCACCGGACGAGCGGCC 484

QY 481 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTGAACAG 540
Db 485 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTGAACAG 544

```


QY	232	CTACGACACATTGAACCTTTTTCAGTGCAAAAAGTAGCGTGTGGC	278
Db	166	CTACTACATCATCAACTTTTTTGCATTACAAAAAAGTTCATTTTTCG	120
RESULT 9	A48542/c		
LOCUS	A48542	Sequence 1 from Patent WO9601320.	DNA linear
DEFINITION	A48542		
ACCESSION	A48542.1	GI:2302312	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Bishop,D., Possee,R. and Ayres,M.		
TITLE	AUTOGRAPHIA CALIFORNICA COMPLETE GENOME SEQUENCE		
JOURNAL	Patent: WO 9601320-A 1 18-JAN-1996;		
COMMENT	NATURAL ENVIRONMENT RES (GB)		
FEATURES	Other publication AU 2897295 960125.		
source	Location/Qualifiers		
	. .133894		
	/organism="unidentified"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32644"		
ORIGIN			
Query Match	18.0%;	Score 101.4; DB 6; Length 133	
Best Local Similarity	75.4%;	Pred. No. 1.2e-20;	
Matches 126;	Conservative 0;	Mismatches 41; Indels	
QY	112	AACCTTTTTCGACTGCAAAAAACACGCTTTTGCACGCGGCCCATACATATA	133
Db	132384	AATTTTTTTCGAAATGCAAAAAAGTTCACITTTTGCCTTGACACTCCATATAC	120
QY	172	CTACGTTTCGTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCT	278
Db	132324	CTACAAATCGTAGACTATTTTATTAGAATAGTCTACACTGTACGATACGCT	132
QY	232	CTACGACACATTGAACCTTTTTCAGTGCAAAAAGTAGCGTGTGGC	278
Db	132264	CTACTACATCATCAACTTTTTTGCATTACAAAAAAGTTCATTTTTCG	132
RESULT 10	L22858/c		
LOCUS	L22858	Sequence 1 from Patent WO9601320.	DNA circular v
DEFINITION	L22858		
ACCESSION	L22858.1	GI:510708	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Ayres,M.D., Howard,S.C., Kuzio,J., Lopez-Ferber,M. an		
TITLE	The complete DNA sequence of Autographa californica n		
JOURNAL	polyhedrosis virus		
MEDLINE	Virology 202 (2), 586-605 (1994)		
PUBMED	94303173		
AUTHORS	8030224		
TITLE	2 (bases 1 to 133894)		
JOURNAL	Kuzio,J.		
FEATURES	Submitted (08-MAR-1999) NCBI, Bethesda, MD 20894, USA		
source	Location/Qualifiers		
	. .133894		
	/organism="Autographa californica nucleopoly		
	/mol_type="genomic DNA"		

repeat_region
/db_xref="taxon:46015"
/clone="C6"
join(133883..133894,1..445)
/standard_name="hri"
/notes="5 copies of 30 bp imperfect palindromic sequence;
the EcoRI site in the first palindrome is at residue 1 of
the linearized genome"
/function="enhancer; replication origin"
/rpt_type=dispersed
503..1009
/gene="Ac-ptp"
/note="synonym: AcOrf-1"
503..1009
/gene="Ac-ptp"
/note="FTP; 19288 Da primary translation product"
/codon_start=1
/product="protein tyrosine phosphatase"
/protein_id="AAA66631.1"
/db_xref="GI:559070"
/translation="MFPARWHNYLCGGVINKDSNLICFKTPRLPELAYVTSEEDVMT
AEQIVKQNPISGAILDITNSKYDGVHFLRAGLLYKKIQVPGQTLPEISIVQBFIDT
VKETKCPGMLGVCHTGINRTYGMVCRYLMHTLGIAPQEAIDRFEKARGHKIERQ
NYVQDLII"
complement(1041..2027)
/gene="Ac-br"
/note="synonym: AcOrf-2"
complement(1041..2027)
/gene="Ac-br"
/note="BRO; 37769 Da primary translation product"
/codon_start=1
/product="baculovirus repeated ORF"
/protein_id="AAA66632.1"
/db_xref="GI:559071"
/translation="MARVIGIEFKPGEDTFNRLVRLERDQQRVFAKDVANSLKYYVC
DKAIRVNDNKSLFEQITQNGFTNSVVKRGDPLYLQPHYVLITKSGVIQIMKS
KLPAIQLQEWLLEEVLPQVLCTGKYDPAIKQREESKQLVKLIATFHTNALQAV
VAQTEELVKKQEFIERIVAIKDQIEAKLOQVMTDLNRMVTFQETMOKKDEIM
QKDAQVTDLVAKVVDLSRAVOYPAOKRHPVLCVTRDGTFTAITGOKTYVENQKH
KENINVANIIVENIRPNTVDWNNATDLQAKRSKRSILVVRKKRNKLIG"
complement(2084..2245)
/gene="Ac-ctx"
/note="synonym: AcOrf-3"
complement(2084..2245)
/gene="Ac-ctx"
/note="CTX; 5590 Da primary translation product"
/codon_start=1
/product="conotoxin-like peptide"
/protein_id="AAA66633.1"
/db_xref="GI:559072"
/translation="MQIKTVLLAFAMFAALNAQHVAACAFTGAVCVHNDCCSGACS
PIFNVCJLPQ"
complement(2295..2750)
/gene="AcOrf-4"
complement(2295..2750)
/gene="AcOrf-4"
/note="17577 Da primary translation product"
/codon_start=1
/product="AcOrf-4 peptide"
/protein_id="AAA66634.1"
/db_xref="GI:559073"
/translation="MKUTYKMASSLKVALRLTRYKENIIPHFDHLRLRLDIMGTK
SEDEVORNTNRNDLISACMOINVRTYMPNATIOMRKPNCIYFRICQYCHLEADVFS
PDDHSVRYLVCAGTPLVIDHPLDVFEGHTEGYNELLEVRVNAAGEL"
2779..3108
/gene="AcOrf-5"
/note="12435 Da primary translation product"
/codon_start=1
/product="AcOrf-5 peptide"
/protein_id="AAA66635.1"
/db_xref="GI:559074"
/translation="MYRTSRINNAVPVAVSQHDYDRDQIKRELNSLRNVHDLCTRSGT

gene
/db_xref="taxon:46015"
/clone="C6"
join(133883..133894,1..445)
/standard_name="hri"
/notes="5 copies of 30 bp imperfect palindromic sequence;
the EcoRI site in the first palindrome is at residue 1 of
the linearized genome"
/function="enhancer; replication origin"
/rpt_type=dispersed
503..1009
/gene="Ac-ptp"
/note="synonym: AcOrf-1"
503..1009
/gene="Ac-ptp"
/note="FTP; 19288 Da primary translation product"
/codon_start=1
/product="protein tyrosine phosphatase"
/protein_id="AAA66631.1"
/db_xref="GI:559070"
/translation="MFPARWHNYLCGGVINKDSNLICFKTPRLPELAYVTSEEDVMT
AEQIVKQNPISGAILDITNSKYDGVHFLRAGLLYKKIQVPGQTLPEISIVQBFIDT
VKETKCPGMLGVCHTGINRTYGMVCRYLMHTLGIAPQEAIDRFEKARGHKIERQ
NYVQDLII"
complement(1041..2027)
/gene="Ac-br"
/note="synonym: AcOrf-2"
complement(1041..2027)
/gene="Ac-br"
/note="BRO; 37769 Da primary translation product"
/codon_start=1
/product="baculovirus repeated ORF"
/protein_id="AAA66632.1"
/db_xref="GI:559071"
/translation="MARVIGIEFKPGEDTFNRLVRLERDQQRVFAKDVANSLKYYVC
DKAIRVNDNKSLFEQITQNGFTNSVVKRGDPLYLQPHYVLITKSGVIQIMKS
KLPAIQLQEWLLEEVLPQVLCTGKYDPAIKQREESKQLVKLIATFHTNALQAV
VAQTEELVKKQEFIERIVAIKDQIEAKLOQVMTDLNRMVTFQETMOKKDEIM
QKDAQVTDLVAKVVDLSRAVOYPAOKRHPVLCVTRDGTFTAITGOKTYVENQKH
KENINVANIIVENIRPNTVDWNNATDLQAKRSKRSILVVRKKRNKLIG"
complement(2084..2245)
/gene="Ac-ctx"
/note="synonym: AcOrf-3"
complement(2084..2245)
/gene="Ac-ctx"
/note="CTX; 5590 Da primary translation product"
/codon_start=1
/product="conotoxin-like peptide"
/protein_id="AAA66633.1"
/db_xref="GI:559072"
/translation="MQIKTVLLAFAMFAALNAQHVAACAFTGAVCVHNDCCSGACS
PIFNVCJLPQ"
complement(2295..2750)
/gene="AcOrf-4"
complement(2295..2750)
/gene="AcOrf-4"
/note="17577 Da primary translation product"
/codon_start=1
/product="AcOrf-4 peptide"
/protein_id="AAA66634.1"
/db_xref="GI:559073"
/translation="MKUTYKMASSLKVALRLTRYKENIIPHFDHLRLRLDIMGTK
SEDEVORNTNRNDLISACMOINVRTYMPNATIOMRKPNCIYFRICQYCHLEADVFS
PDDHSVRYLVCAGTPLVIDHPLDVFEGHTEGYNELLEVRVNAAGEL"
2779..3108
/gene="AcOrf-5"
/note="12435 Da primary translation product"
/codon_start=1
/product="AcOrf-5 peptide"
/protein_id="AAA66635.1"
/db_xref="GI:559074"
/translation="MYRTSRINNAVPVAVSQHDYDRDQIKRELNSLRNVHDLCTRSGT

SDFCNKFLRSDDMTVPVTTITPKRTADYKITEYVGDVKTIKPSNRPLVESGPLYVREA
KXGECIV"
3089..3721
/gene="Ac-lef2"
/note="synonym: AcOrf-6"
3089..3721
/gene="Ac-lef2"
/note="LEF2; 23926 Da primary translation product"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAA66636.1"
/db_xref="GI:559075"
/translation="MANASYNVMSPLIRASCLDKKATVLDIPDDFIDKLTLPYTVFY
NGGVLVKISGLRLYMLLTAPPTNEIKNSNFKRSKENICMKECEVKKNVVDMLNKK
INMPPCIKKILNDLKNNVPRGMYRKRPFILNCYIANVVSACAKENCCLIKALHTFYN
HDSKCGEVMHLIKSQDVKYPCNKQKMTVDKLCPPAGNCKGLNPLCNY"
complement(3759..4364)
/gene="AC-ORF603"
/note="synonym: AcOrf-7"
complement(3759..4364)
/gene="AC-ORF603"
/note="ORF603 peptide; 23612 Da primary translation
product"
/codon_start=1
/protein_id="AAA66637.1"
/db_xref="GI:559076"
/translation="MAVIFNNKQLLADNSIEKGELFPLNGSYNLSYVNPVLLKNG
VIELEAAAYAGNILEYKTDPKFIDYINLIKATHSELPELNSVTVNRYRKTMRSGTIIH
PIKKDIYIDNKKFTLYDYGYDNNYVNFYEKNEKEKEEEDDDKASSLCENKII
LSQINCSFENDFKYLYSDIYAFSIIIDNTNLVAFGLYR"
4520..5257
/gene="AC-PH"
/note="synonym: AcOrf-8"
4520..5257
/gene="AC-PH"
/note="polyhedrin; 28642 Da primary translation product"
/codon_start=1
/product="major occlusion body protein"
/protein_id="AAA66638.1"
/db_xref="GI:559077"
/translation="MPDYSRYPTIGRTYVVDNKYKNGLAVIKNAKAKKKHFAEHEIEE
ATLDLDNLYVAEDPFLGPKGKQLTLFIRNNYKPDTKMLVGVKWKGEFYETWTRF
MEDSPFIIVNDQEVMDVFLVMVTRPNRCYKFLAQHALRCDPDYVPHDVIRI VEPWS
VGSNNEYRISLAKKGGCCPIMNLHSEYTNSEFQIDRVWENFYKPIVYICTDSAEIE
EILLESVLVKVKEFAFDAPLETGPAY"
complement(5287..6918)
/gene="AC-ORF1629"
/note="synonym: AcOrf-9; Ac-61K; Ac-vp78"
complement(5287..6918)
/gene="AC-ORF1629"
/note="61K; vp78; required for virus replication; 60713 Da
primary translation product"
/codon_start=1
/product="viral capsid associated protein"
/protein_id="AAA66639.1"
/db_xref="GI:559078"
/translation="MTNRRYVSQSYLFNNRNKKIDAOEFERVDTAEAQILIKNIYD
NTVNLNRDLNLKLANDVFNKAYMVDSEVSRHYNVAVVMKRLVIGVRDLSLRQ
SLYNTIAYIERLNLIGTNDSEITMLIADFYDLNSYNIELPPQALPRSRRSVQRP
AAPAPVTIVREOTKPKQIIIPAPPPPPSPVNIAPPPPPSMSELPPAPMPTEP
QAAPLDDRQLQLEAIENKRNTRLRPVKPTAPETSTIVEVTVLPKETPEPKPPSA
SPRPPPPPPPPAPAPPMVDLSSAPPPLVDLPSEMLPPAPPSLSNLSSELKSGT
VRLPKAQKRSQSEIIPKSSNTLIADVLADTINRRRVAMAKSSSEATSNDSGWDGDDN
RPNKANTPDYKVOALENVFTSSQLYINDSDERTKAHNILNDVPELLQNKQTQNIIDK
ARLLQLDLASVALSENPLDSPGISEKQPUFETRNRLFYKSIDLIFKFKYKOAENH
LIFALTYHPKDYKFNELLYKVVQVLSVNOQRTSSA"
6917..7735
/gene="Ac-pk-1"
/note="synonym: AcOrf-10"
6917..7735
/gene="Ac-pk-1"
/note="PK1; 31978 Da primary translation product"
/codon_start=1

```
/product="protein kinase"
/protein_id="AAA66640.1"
/db_xref="GI:559079"
/translation="NATNATLOTLOVQYVENCNKVTRYKLIINGRFGKISILSHKPTS
KLYLOKTSANFNADIKVHQLMSDHPNFIKLVNHSINNQVIMDYIDCPDLFET
LOIKGELSYQLVSNIRIOLCEALNDLKHNFHNDIKLVNLYFEALDRVYVCDYGLC
KHNSLSVHDGTLEYFSPEKIRHTMYHVSFDWYAVGLTYKLLTGRRHPFKSDEML
DLNSMKRQOQYDGLVGLKVRNVNARDFYVCLTRYNDCLRLTYKQIIKHFFLS"
7477. .7864
/standard_names="hrla"
/note="2 Copies of 30 bp imperfect palindromic sequence"
/function="enhancer; replication origin"
/rpt_type="dispersed"
complement(7899. .8921)
/gene="AcOrf-11"
complement(7899. .8921)
/gene="AcOrf-11"
/note="AcOrf-11"
/note="40093 Da primary translation product"
/codon_start=1
/product="AcOrf-11 peptide"
/protein_id="AAA66641.1"
/db_xref="GI:559080"
/translation="MSLAAKLIIYNYAKYNEVHDVCGESYHHYRIYQVYLSSEYVNG
MSCIERDVTARRLSKSGCTFDEAVKMDAGDSIKLSHWFSTETMGIDNVRVLE
QIDAVPVSVRVQNGQIFSLNNPFEISQDMLDCLQIILGRFPEYFMNGKLLRIANV
FNPNNVGVWYVYVYVYHRIYRMYSPAEVLPRLSEAVKKTIRLRSKYDRLHV
FNSNCPRIAYEMGRFCIGKEHPSKHLSCHMILFOLYLRGKTQBEKSPCYRVIK
DFGROCKDYVKNLKDVFLLHSHMSDKDKNSMLDLCVMBCEEIDVDFYVIFESFL
NNKHQ"
Query Match 18.0%; Score 101.4; DB 14; Length 133894;
Best Local Similarity 75.4%; Pred. No. 1.2e-20;
Matches 126; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
Qy 112 AACTTTTGGCACTGCAAAAACAGCTTTTGCAGCGGGCCCATACATAGTACAAACT 171
Db 132384 AATTTTTTGGCAATGCAAAAAGTTCACTTTTGCCTGACATCCCATATACATCAACT 132325
Qy 172 CTACGTTTCTGATAGTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACA 231
Db 132324 CTACAAATCGTAGACTATTTTATTAGATAGTCTACATGTCAGTACGATCCCATATA 132265
Qy 232 CTACACACATGAACTTTTGGAGTGCAGTCAAAAAGTACGTTCCGC 278
Db 132264 CTACTACATCACTTTTGGCATTACATAAAAAGTTCACTTTTGC 132218
RESULT 11
AY327402 131158 bp DNA circular VRL 08-OCT-2003
LOCUS Choristoneura fumiferana defective nucleopolyhedrovirus complete
DEFINITION
ACCESSION AY327402 AF068194 U10476 U23422 U78194
VERSION AY327402.1 GI:37499238
KEYWORDS
SOURCE
ORGANISM
Choristoneura fumiferana defective nucleopolyhedrovirus (CfDEFNPV)
Choristoneura fumiferana defective nucleopolyhedrovirus
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus
REFERENCE
1 (bases 11500 to 13512)
Barrett,J.W., Krell,P.J. and Arif,B.M.
Characterization, sequencing and phylogeny of the ecdysteroid
UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis
viruses isolated from Choristoneura fumiferana
J. Gen. Virol. 76 (Pt 10), 2447-2456 (1995)
96030854
MEDLINE
PUBMED 7595348
REFERENCE 2 (bases 13243 to 14241)
Barrett,J.W., Lauzon,H.A., Mercuri,P.S., Krell,P.J., Sohi,S.S. and
Arif,B.M.
The putative LEF-1 proteins from two distinct Choristoneura
fumiferana multiple nucleopolyhedroviruses share domain homology to
eukaryotic primases
```

```
JOURNAL MEDLINE
PUBMED 97187920
3 (bases 80268 to 81693)
REFERENCE
AUTHORS Li X., Lauzon,H.A., Sohi,S.S., Palli,S.R., Retnakaran,A. and
Arif,B.M.
TITLE Molecular analysis of the p48 gene of Choristoneura fumiferana
multicapsid nucleopolyhedroviruses cfMNPV and CfDEFNPV
J. Gen. Virol. 80 (Pt 7), 1833-1840 (1999)
99350016
MEDLINE 10423153
PUBMED 4 (bases 1 to 131158)
REFERENCE
AUTHORS Li,X., Barrett,J., Pang,A., Klose,R.J., Krell,P.J. and Arif,B.M.
TITLE Characterization of an overexpressed spindie protein during a
baculovirus infection
Virology 268 (1), 56-67 (2000)
20149221
PUBMED 10683327
REFERENCE
AUTHORS Lauzon,H.A.M., Jamieson,P.B., Krell,P.J. and Arif,B.M.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2003) Molecular Virology, Great Lakes Forestry
Centre, 1219 Queen St. E., Sault Ste. Marie, Ontario P6A 2E5,
Canada
COMMENT On or before Oct 8, 2003 this sequence version replaced gi:1117788,
gi:2501770, gi:1754838, gi:4092491.
FEATURES
source
1. 131158
/organism="Choristoneura fumiferana defective
nucleopolyhedrovirus"
/mol_type="genomic DNA"
/db_xref="taxon:74660"
join(129349. .131158,1. .20)
/note="ORF 149: 1629 capsid Op2/Ac9"
/codon_start=1
/product="viral capsid associated protein"
/protein_id="AAQ91783.1"
/db_xref="GI:37499384"
/translation="MDRQYQSVKGLNNNAHAIDAADAPQVLVAGPAHNLKRLGFG
RVKLNATLIDLKLAENIYADTAYMQVNTIESSRHFPATLMRNLNLLNVDQEHFRN
TSLKVARINLESDANDVEITVLSGDVEEYSRYATROYVSTDTLPPPTLPI
PETQAFVSPPPQAQFVAPSPPAQAFVAPSPPAQAFVAPSPQPTFVAPSPQPT
FVAPSPQPTFVAPSPQPTFVAPSPQPTFVAPSPQPTFVAPSPQPTFVAPSPQPT
NGVNLTKLPVPPKPAHLSPNFMFVGDVGTNTPPPNGTSPPQGVNVPVPPVAP
PLNVPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
PPSLRGDMLNQRITGATLRKTGLENDLDDMMSEPRKATRSALFDQIKMGATLKAGQVE
RSDVATSESTSGFDESADTRADKASKSELKHAHLNFAKDKSLYNLOKVNSELTKI
DENVPPLKRSPTAEVNEKANGLYLFRQHVTLPPKALDAQAPALYAADAPQYVQ
IEDLLFAGRYDDARAFIOAVDAPEDMKLKKFELTVANQLSTRGQ"
complement(1. .738)
/note="ORF 1; ph Op3/Ac8; major occlusion body protein"
/codon_start=1
/product="polyhedrin"
/protein_id="AAQ91696.1"
/db_xref="GI:37499297"
/translation="MPNYSRPTTIGRTVYDKNYKNGLSVINKAKRKKHLEHEEDE
KHLDPHYVMVAEDPFLGPKGNKQLTKFKEIRNVKPDPTMKLIWNMSGFEFLRETWTRF
VEDSPFVNDQEVMDVFLVNLNRPNRCYKFLAQHALRWDCCVPHVEVIRIVPSY
VGMNVEYRISLAKKGGCPMNIHSEYNTSNSESVFNVRVWENFYKPIVIGTDSGEE
EILLEVSLVFKVEFADAPLFTGPAY"
complement(828. .1220)
/note="ORF 2; Op5"
/codon_start=1
/product="unknown"
/protein_id="AAQ91742.1"
/db_xref="GI:37499343"
/translation="MADSTNRILRFLYRWSKGTGASLENQADLECLVDLRFVGAHMN
KQITCKKKQAEERAEASALKVELAADRAMQAKAAPYCADDDGRWSTLSQOOLE
IDRKEIVDRYIKLQKQDCILNNKNAF"
complement(1247. .1858)
/note="ORF 3; lef-2 Op6/Ac6"
/codon_start=1
/product="late expression factor 2"
```



```
/protein_id="AAQ91716.1"
/db_xref="GI:37499317"
/translation="MGSAMNPAAGVSQSLNSEKYLIDPNDPFCVGLALSPTCTVFKQGL
FMENSGRLRALLTAPTEPKRAVLHRSKRNVCCKACAGSVNLAKALNSLRPLCM
VTHAEELNNAAPRGMYRRKRFETCYLGNVSVCTCKSACLIGLHFFYKMDPKCVG
EVHLLKAEVDYKSPSCAKNKTYNKLCPKAGMCKGNPICNF"
CDS
/notes="ORF 4; Op7/Ac5"
/codon_start=1
/product="unknown"
/protein_id="AAQ91760.1"
/db_xref="GI:37499361"
/translation="MNRPMRMTANAPVSVNHDYDRQIRDLNLSRRSVHELCTRET
TGPDNRILELSSSSNDVITKSPVIKNSAATVQSALVCDV"
CDS
2171..2608
/notes="ORF 5; Op8/Ac4"
/codon_start=1
/product="unknown"
/protein_id="AAQ91737.1"
/db_xref="GI:37499338"
/translation="MASARFTKSLRVTCQFENIVADITHLARLRALIDGKTHTDV
HRFGLRNLALVEACMAVNOVVPDGTIRLNPKTIYFRVCQKCHAVDPAPDDHSI
ARYLCAACGMVLVLDHPLDVGDTBEGVNELIEVQRINAGDGL"
CDS
2647..3657
/notes="ORF 6; bro Ac2"
/codon_start=1
/product="baculovirus repeated ORF"
/protein_id="AAQ91674.1"
/db_xref="GI:37499275"
/translation="MAQIKIQGQKFGEDVFTLRVYLRDRIKFKVAKDVASSLKYNICD
KAIRTHVDDKYKTFEQTILQGSSTNLVKRGDPLYLQPHTVLITKSGVIQLIMKSK
LPYAEIQEWLLEVIQVLCTGYKNPAIKQQQENKQLINKLVKTFSDHTNTLQAL
LQKTELKQKEFTIERVATKDKQLEAKDLQVTRWTDLRMYNGFOETWQKDEIMQ
QKQADIVLVAKVIDLSGRAVQYFEDERKGPVLCVTRDGTFTTAIGCQKAYVHNQKUK
RNLGSAADIVATETPRPVDWNNATHRLAAKSKRTISFDSEQDAQGFATRIKOLLNN
FD"
complement(3696..4168)
/notes="ORF 7; ptp-2 Op9"
/codon_start=1
/product="protein tyrosine phosphatase 2"
/protein_id="AAQ91731.1"
/db_xref="GI:37499332"
/translation="MYDANQIDENLFVGGYGYNNAMQLFKKHDIBSVLISDSVG
PIROALGPAGYHIHVCEDEPTCMALINAMDALYDIERRINEGKILIHCHAGVSR
SATLAVYYMKKWQSVSEKALRFVNNKRNVALSDHFVFLSSRCTYRFVNNKLIHVG
"
complement(4155..4676)
/notes="ORF 8; ptp-1 Op10/Ac1"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAQ91726.1"
/db_xref="GI:37499327"
/translation="MFGRMHNYTTCGVIEGTNLICFKVPLKOEVEFYVNDSDRWIT
VSNLVNQRLAGVADIDTLTRYDYGAKVRDAGVLYKKIRVPQGEVNEIDIVQEFDS
VQEFSAKCPGMLIGVHCTHGLNRTGYLCVRYIIDKLHVSPTDAIARFEAAARGHKIERR
NYLEDLLKRHVR"
CDS
4935..5921
/notes="ORF 9; Op11/Ac11"
/codon_start=1
/product="unknown"
/protein_id="AAQ91675.1"
/db_xref="GI:37499276"
/translation="MSLCSQLLVAYGYEYNNPHERYSESVHLVRIVNEYILINSYANE
TSCVRRDIETARRLNRELSFDDARRQLNLEELAKRLATWYHTGEMKSPCFSEIQSULA
EINCHAPLETGASRQVAFPAIDSLFPNTDSIQVMGFMHFVAGYLAUTHVAVF
DPTIKLDGWYNTKFCVITYYRIIRGTVPABELITRLQNVTYKIPKPYDESNNALWG
DVGRCFGIGKDHFAQHKMSVYIFQYMRGEVYADERFPCFSVIKDFGRCKETYK
DIQVQDILHSHMDTKQKNALFDLLCCNNASDIDVDYDYIVKKFYNIAY"
6450..6723
repeat_region
/notes="4 copies of a 30 bp imperfect palindromic sequence;
location similar to Ac hria"
/function="enhancer; replication origin"
complement(6723..7871)
CDS
```

```
/note="ORF 10; Op20/Ac20"
/codon_start=1
/product="unknown"
/protein_id="AAQ91663.1"
/db_xref="GI:37499264"
/translation="MYRILIVFLFALLYIVVMPFYQAYQHIOAQHDIYNDTLTDRMD
DPGASVFFVGGEGTYVNGHKLAVGGYCTTNSVPRDCNRETSVVLSLNOQWTCIAEDP
YIESVMRRHYVPMELPSVQFDNLGTLAGETLKCMSVLYVTEILDLPFDCQVCE
RYFAGTSMTQLAGRHFDRLILPQSDRNVLFDRLLGREVNVAINTFRSDELEDG
TRRFEMKRNARDNNNLMFVNLPLDLECPNCTNVNYVHTSVRPNFETGECCGDEA
VTRVRHVPGDRSSVCASIVDGLDITTTASHRFVECVTYTISGNFNNKLLCPSDTF
DSNTDAFAFEVPCGYSPLSGNGLDEPHTREFLDTRSRIRYNDVRGLIN"
7898..8863
/notes="ORF 11; arif Op19/Ac21"
/codon_start=1
/product="actin rearrangement inducing factor"
/protein_id="AAQ91678.1"
/db_xref="GI:37499279"
/translation="MLAQINYLQLVLAALYTITLIAFVFLSMGTINYKYAFLELN
DNHGSVNLGLTAFLLGPCVFTTTWAIYKFLCYKRAEMHSNFYKTTIISLAHVMA
LVKWTLPVQPOIYKNGHVPVLDARYDYDRSLCHSNIVSDTYEVHDTNLAJFDEN
CVTRHDFVKKCVGCRMEVRHDEPTVFNQNOCALIMWMNTAVLOFWNNYVQRKEMRYK
PTPVKTLYFESAPLKEQDTADEEEQSSFRMLEIIESEPRVQFQFPSSSLDRUSSPP
PIVQSSSPNSPDSGIDYDIPQPFYSVPNKVCKYLCTHATLCA"
complement(9013..9327)
/notes="ORF 12; Op18/Ac19"
/codon_start=1
/product="unknown"
/protein_id="AAQ91753.1"
/db_xref="GI:37499354"
Query Match 17.7%; Score 100; DB 14; Length 131158;
Best Local Similarity 80.6%; Pred. No. 3.3e-20;
Matches 129; Conservative 0; Mismatches 30; Indels 1; Gaps 1;
Qy 112 AACTTTTTCGACTGCAAAAACACGCTTTTGCACGCGGCCCATACATA-GTCAAAAC 170
Db 125537 AACATTATTGAGTGCAAAAAGTTTCACTTTATATCGACCTTATACACATCAAC 125596
Qy 171 TCTACGTTTCGTAGACTATTTTACATAAATAGTCTACACGTTGTATACGTCGAATAC 230
Db 125597 TCTACGTTTCGTAGACTATTTAAGTGAATAGTCTACACTGTGTATACGTCCTCAATAC 125656
Qy 231 ACTACACACATTCGAACCTTTTTCAGTCGCAAAAAGTAC 270
Db 125657 ACTACACACATCGAATTTTTCGACTGCAATAAAGTTC 125696
RESULT 12
BD187790/c 28413 bp DNA linear PAT 17-JUL-2003
LOCUS A virus in which a gene for controlling an insect behavior is
DEFINITION deficient.
ACCESSION BD187790
VERSION BD187790.1 GI:32997529
KEYWORDS JP 2003024062-A/3.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 28413)
AUTHORS Kang,W., Imai,N., Gomi,S. and Matsumoto,S.
TITLE A virus in which a gene for controlling an insect behavior is
JOURNAL Patent: JP 2003024062-A 3 28-JAN-2003;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
COMMENT OS Bombyx mori nucleopolyhedrovirus
PS JP 2003024062-A/3
PF 28-JAN-2003
PI 10-JUL-2001 JP 2001209305
PI2N15/09,A01K67/033,C12N7/04
CC From 100,001 to 128,413 of Bombyx genome DNA
FH Key location/Qualifiers
FT source 1..28413
```

FEATURES	source	FT
ORIGIN		<p>Location/Qualifiers</p> <p>1..28413</p> <p>/organism="Bombyx mori nucleopolyhedrovirus"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:32644"</p>
Query Match		16.6%; Score 93.4; DB 6; Length 28413;
Best Local Similarity		72.5%; Pred. NO. 3.2e-18;
Matches 121;	Conservative	0; Mismatches 46; Indels 0; Gaps 0;
QY	112	AACATTTTTCGACTCGAAGAAAAACAGCTTTTGCACGGGGCCCATACATAGTACAAACT 171
DB	22231	AAATTTTTTTCGCAATACAAAAAGTTCCGCTATGTTTGACATATAATATACAGTACGAACT 22172
QY	172	CTACGTTTTCGTGACATATTTTACATAAATAGTCTACACCGTTGTATAGCTCCAAATACA 231
DB	21171	CTCAAAATCGTAGACATATTTTATTAGAAATAGTCTACATGTACTATAGCTCTCAATATA 22112
QY	232	CTACCACACATTTGAACCTTTTTCGAGTGCAGAAAAAGTACGTCTCGGC 278
DB	22111	CTACTACATCATCACTTTTTCGCAATACAAAAAGTTTCATTTTTCG 22065
RESULT 13		
NPHT3COMP/c		
LOCUS	NPHT3COMP	128413 bp DNA circular VRL 05-MAY-1999
DEFINITION	Bombyx mori nuclear polyhedrosis virus isolate T3, complete genome.	
ACCESSION	L33180	
VERSION	L33180.1	GI:3745835
KEYWORDS		
SOURCE	Bombyx mori nucleopolyhedrovirus	
ORGANISM	Bombyx mori nucleopolyhedrovirus	
REFERENCE	Viruses; dsDNA viruses, no RNA stage; Baculoviridae;	
AUTHORS	Nucleopolyhedrovirus.	
TITLE	1 (bases 1 to 128413)	
JOURNAL	Kamita, S.G. and Maeda, S.	
MEDLINE	Sequencing of the putative DNA helicase-encoding gene of the Bombyx	
PUBMED	mori nuclear polyhedrosis virus and fine-mapping of a region	
REFERENCE	involved in host range expansion	
AUTHORS	Gene 190 (1), 173-179 (1997)	
TITLE	9185864	
JOURNAL	2 (bases 1 to 128413)	
MEDLINE	Gomi, S., Majima, K. and Maeda, S.	
PUBMED	Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus	
REFERENCE	J. Gen. Virol. 80 (Pt 5), 1323-1337 (1999)	
AUTHORS	99281911	
TITLE	10355780	
JOURNAL	3 (bases 1 to 128413)	
MEDLINE	Maeda, S.	
PUBMED	Direct Submission	
REFERENCE	Submitted (29-MAY-1994) Department of Entomology, University of	
AUTHORS	California, Davis, CA 95616, USA	
TITLE	4 (bases 1 to 128413)	
JOURNAL	Gomi, S.	
MEDLINE	Direct Submission	
PUBMED	Submitted (14-OCT-1998) Laboratory of Molecular Entomology and	
REFERENCE	Baculovirology, The Institute of Physical and Chemical Research	
AUTHORS	(RIKEN), Hirose, T., Wako 351-0198, Japan	
TITLE	Sequence update by submitter	
JOURNAL	On Oct 14, 1998 this sequence version replaced gi:1196668.	
MEDLINE	Location/Qualifiers	
PUBMED	1..128413	
REFERENCE	/organism="Bombyx mori nucleopolyhedrovirus"	
AUTHORS	/mol_type="genomic DNA"	
TITLE	/isolate="T3"	
JOURNAL	/db_xref="taxon:271108"	
MEDLINE	1..7738	
PUBMED	/notes="28817 Da primary translation product"	
REFERENCE	/codon_start=1	
AUTHORS	/product="Polyhedrin=PH=AcMNPV orf8"	
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
MEDLINE		
PUBMED		

/codon_start=1
/product="LEF-1=AcMNPV orf14"
/protein_id="AAC63688.1"
/db_xref="GI:3745841"
/translation="MLLCNYYTQKRVDMMDAIAVNDYRKRKYAFMTVFNARWIHADKYFPT
AAQYISYIQNVKSDVHVPLDDGGREVVADYKYNVDEHDLMLKIYIGATAFLIF
YTEVSRVMYVYGNRSHLMLKFTDKFKITQAQNVVRHYKAPFKPAKLDSDCIQPGS
FAHCVREAVRLYIPHMDSNLDAITLOYPDVPDRDIFCNVKNQIRAPYSYNYKGTKEF
RCITKELLDLKQCYPGYGGCGPVTITTTTTTTPPSPPKIGSIQTTTKSTT"
6407.7927
/gene="egt"
6407.7927
/gene="egt"
/note="57044 Da primary translation product"
/codon_start=1
/product="UDP-Glucosyl Transferase=EGT=AcMNPV orf15"
/protein_id="AAC63689.1"
/db_xref="GI:3745842"
/translation="WTILCWLALLSTLTAVNAVNI LAVFPTPAYSHIYVKYIEALA
EKCHNVTVKPKLPAYSTKTCVCGNI TEVNSDMSVKQYKLVTVNSAMFRKGVSDTDT
VTAANYLGLTEMFQDQNDINVRNLIANQTFDLVVVEAFADYALVFGHLYDPAPVIQ
IAPGGAENFDTIGAVARHVPVHPNIRNNFDDTKANLMTMLELYKEFKILANMSNA
LLKQFGPDTTIEELNQRQLLLNLHPIDNNRPSPSVQYLGSGIHLVKSAPLTK
LSPVIDAMNKSAGAIYVSGSSIDTSFANEFYMLINTFKALDNYITLWKIDDBV
VKNITLPAVITQWNRVLRHKMAAFITQGGLOSSDEALPAGIPMVLPMWQDQ
FYAHKLQQLGVARALDVTVSSDQLLLATNDVLFNASTYKKGMAELVALINNDKATF
PPLDKAIFTEVRVRYRHDISRRLYSLKTAANVPYSYNYKSVLSIVMNHIAHF"
7940.8101
/gene="Orf_7a"
7940.8101
/gene="Orf_7a"
/note="6233 Da primary translation product"
/codon_start=1
/product="Orf_7a"
/protein_id="AAC63690.1"
/db_xref="GI:3745843"
/translation="MLFIIIVLVCATAHFPAINAHQEPQSLQLQDVQRKHNE
FCSHALMCQ"
8067.8756
/gene="bv/odv-e26"
8067.8756
/gene="bv/odv-e26"
/note="26201 Da primary translation product"
/codon_start=1
/product="BV/ODV-E26=AcMNPV orf16"
/protein_id="AAC63691.1"
/db_xref="GI:3745844"
/translation="MNSVHTRLCASSNQFAPFKRQLAVPVGVSNSLTHITSTVAS
VIPKNYQEKROKI CHIISSRLNTHLNFKIQSVHKKLLHLQNLRLKNEIIAELVRK
LESAQKTTTHRNISSKPAHWKYFGVVRCDNTIRTIIGNEKFRRLAELCTLYNA
EYVFCQARADGDKRQALASLLTAAFGPRVIVYENSRPEINPDEIASGKRLIHL
QDESQSINAY"
8725.9357
/gene="Orf_9"
8725.9357
/gene="Orf_9"
/note="24057 Da primary translation product"
/codon_start=1
/product="AcMNPV orf17"
/protein_id="AAC63692.1"
/db_xref="GI:3745845"
/translation="MNLKVIITLNLKGEPPNCVERITIMPCSLMDTIELNVRCS
PFAKFKVLIVDGFDSAYIOATFCSISDSVTIUNKNEKVMDFVRPDDEGTMWY
VIGLVSVDAAVARVKYKDMVNSIENQOTMLKVFINEANVYNKNMMLKGLIYNNNNE
SVLVNNVKPIKVDKNVNTYKKNVTKWVPVAVNYFTGRQLLTILFIFPKX"
complement(9387.10457)
/gene="Orf_10"
complement(9387.10457)
/gene="Orf_10"
/note="41536 Da primary translation product"
/codon_start=1
/product="AcMNPV orf18"
/protein_id="AAC63693.1"

gene
CDS

gene
CDS

gene
CDS

gene
CDS

/db_xref="GI:3745846"
/translation="MERLNQNLNGLVLPYITTKDIEDRLPDKIVAKAKLAFIKDCPEA
VVCENGFLVLTGGAATVCHIDDDYKSAUKCIDFDYIGLCSKMKFCNLTQINLQKQVD
QVBAEDLVLRQIYMSNPVLMVKCYQAYRLNGQIDLHLNRHIKCIKTQYNDEFDLV
REYALQIDITSAYGVDEYTDNCVKITTPAPLSFNFFVFNVRIMKRPFNADRCIKNPSLLG
NFYHVLVSSIQRVNLQMLCLDKDITFNKFDYKIKRNLHLKRLFANLPAESYNSCVN
DLTDMVLYKQNETITFNVKKIIDISGPDGLGCRKLMRIYLTNTDTFRQQLPAYLTHYVN
YPHKNCLDQNMKRFMSCIFSLY"
Query Match 16.6%; Score 93.4; DB 14; Length 128413;
Best Local Similarity 72.5%; Pred. No. 4.5e-18;
Matches 121; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Qy 112 AACTTTTTCCTGCAAAAAACACCGCTTTTGCCACGGGGCCCATACATAGTACAAACT 171
Db 122331 AATTTTTTTCATACAAAAAGTTTCGCCCTATGTTTGCATATATAATATACATAGTACGAACT 122172
Qy 172 CTAGCTTTCGTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAATACA 231
Db 122171 CTACAAATCTGACTATTTTATTAGTAATAGTCTACACTGTACTATACGCTCTCAATATA 122112
Qy 232 CTACCACACATTTGAACTTTTTCGAGTGCAGTGCAAAAAGTACGTGTGCGG 278
Db 122111 CTACTACACTATCAACTTTTTTGCATTACAAAAAGTTCATTTTTC 122065
RESULT 14
AF512031/c
LOCUS AF512031 129609 bp DNA circular VRL 27-MAY-2004
DEFINITION Choristoneura fumiferana MNPV polyhedrin, complete genome.
ACCESSION AF512031 AF177329 S78506 S81690 U10441 U18677 U26676 U26734 U53854
US7401 U59008 U70432 U72240 X65395 S46001
AF512031.2 GI:47157118
VERSION
KEYWORDS
SOURCE
ORGANISM
Choristoneura fumiferana MNPV
Choristoneura fumiferana MNPV
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE
AUTHORS
TITLE
1 (bases 1 to 129609)
Lee,H.Y., Arif,B., Dobos,P. and Krell,P.
Identification of bent DNA and ARS fragments in the genome of
Choristoneura fumiferana nuclear polyhedrosis virus
Virus Res. 24 (3), 249-264 (1992)
JOURNAL
MEDLINE
PUBMED
93033705
1413988
REFERENCE
AUTHORS
TITLE
2 (bases 1 to 129609)
Xie,W.D., Arif,B., Dobos,P. and Krell,P.J.
Identification and analysis of a putative origin of DNA replication
in the Choristoneura fumiferana multinucleocapsid nuclear
polyhedrosis virus genome
Virology 209 (2), 409-419 (1995)
JOURNAL
MEDLINE
PUBMED
95297142
7778276
REFERENCE
AUTHORS
TITLE
3 (bases 1 to 129609)
Liu,J. and Carstens,E.B.
Identification, localization, transcription, and sequence analysis
of the Choristoneura fumiferana nuclear polyhedrosis virus DNA
polymerase gene
Virology 209 (2), 538-549 (1995)
JOURNAL
MEDLINE
PUBMED
95297155
7778286
REFERENCE
AUTHORS
TITLE
4 (bases 1 to 129609)
Barrett,J.M., Krell,P.J. and Arif,B.M.
Characterization, sequencing and phylogeny of the ecodysteroi
UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis
viruses isolated from Choristoneura fumiferana
J. Gen. Virol. 76 (Pt 10), 2447-2456 (1995)
JOURNAL
MEDLINE
PUBMED
96030854
7595348
REFERENCE
AUTHORS
TITLE
5 (bases 1 to 129609)
Qiu,W., Liu,J.J. and Carstens,E.B.
Studies of Choristoneura fumiferana nuclear polyhedrosis virus gene
expression in insect cells

gene
CDS

gene
CDS

gene
CDS

gene
CDS

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	Virology 217 (2), 564-572 (1996) 96183379 8610448 6 (bases 1 to 129609) Liu, J.J. and Carstens, E.B. Identification, molecular cloning, and transcription analysis of the Choristoneura fumerana nuclear polyhedrosis virus spindle-like protein gene Virology 223 (2), 396-400 (1996) 96400202 8806578 7 (bases 1 to 129609) Lapointe, R., Back, D.W., Ding, Q. and Carstens, E.B. Identification and molecular characterization of the Choristoneura fumerana multicapsid nucleopolyhedrovirus genomic region encoding the regulatory genes pkp, p47, lef-12, and gta Virology 271 (1), 109-121 (2000) 20276145 10814576 8 (bases 1 to 129609) Carstens, E.B., Liu, J.J. and Dominy, C.N. Identification and molecular characterization of the baculovirus CfMNPV early genes: ie-1, ie-2 and p638 Virus Res. 83 (1-2), 13-30 (2002) 21894555 11864738 9 (bases 1 to 129609) de Jong, J.G., Dominy, C.N., Lauzon, H.A., Arif, B.M., Carstens, E.B. and Krall, P.J. Complete Genome of Choristoneura fumerana Multiple Nucleopolyhedrovirus Unpublished 10 (bases 1 to 129609) de Jong, J.G., Dominy, C.N., Lauzon, H.A., Arif, B.M., Carstens, E.B. and Krall, P.J. Direct Submission Submitted (13-MAY-2002) Department of Microbiology, University of Guelph, Chemistry and Microbiology Building, Guelph, ON N1G 2W1, Canada 11 (bases 1 to 129609) de Jong, J.G., Dominy, C.N., Lauzon, H.A., Arif, B.M., Carstens, E.B. and Krall, P.J. Direct Submission Submitted (13-MAY-2004) Department of Microbiology, University of Guelph, Chemistry and Microbiology Building, Guelph, ON N1G 2W1, Canada Sequence update by submitter On May 13, 2004 this sequence version replaced gi:30269978. Location/Qualifiers 1. .129609 /organism="Choristoneura fumerana MNPV" /mol_type="genomic DNA" /db_xref="taxon:208973" /country="Ireland" complement(1. .735) /note="ORF1; Ac8/Op3" /codon_start=1 /product="polyhedrin" /protein_id="AAP29795.1" /db_xref="GI:30269979" /translation="MPDYSYRTIGTYYVYDKNYKNGSVIKRKKHLLHEDEK HLDPLDHYNDVDFPGKNQKLTLPKEIRNVKPTMKLIYNVSGKFLRETWTRFY EDSFTIVNDQEVDFLVVNMETPRNRCYKFLRQALRWDCDYVPHEVIRIVEPSYV GNNEVYRLAKKGGCGCPIMNIHAETVNTSFNFVRNVIWENFYKPIVIGTDSGEEE MLIEVSLVPKVEFAPDAPLFTGPAY" complement(838. .1449) /note="ORF2; Op5" /codon_start=1 /product="unknown" /protein_id="AAP29796.1" /db_xref="GI:30269980" /translation="MQGGVDFDRAAALLQDQCGMCGRGDAFVDQSRHIIQAVQLRENE GHKVVPKSKHVQGLESLQLLVGVSKHGKMAASDQKILELPYRWSSQTGCEMLDDEK
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	DLHCLYDLERFVGHLNKRCDAAKKNKCAKAAALKRVOISADRHAAEAAAAPIADDD GQWSKLTSAQLEIARERXDIYDRIYQLQKQDLIKMDGLKKQ" complement(1251. .1859) /note="ORF3; lef2; Ac6/Op6" /codon_start=1 /product="late expression factor 2" /protein_id="AAP29797.1" /db_xref="GI:30269981" /translation="MDQVWNPAAAGAGSVKPKETYLIDPNDVFGMLELTPYTVFERGLF IRMSGMRLLALLAAPKPKVPTTRFPORSKRNVCLEKADGQPSQSLTKVLTARMNPLC MSKIMADIGSAPRGNMVYKREFEENCYLANVLCTCKKACALIGALLHFRMDAKCVGE VTHLLIKAEITYKPSNCAKMAVTKLCPKANCKGLNFCNY" complement(1862. .2092) /note="ORF4; Ac5/Op7" /codon_start=1 /product="unknown" /protein_id="AAP29798.1" /db_xref="GI:30269982" /translation="MIRPTRNAAAYADYDREQLRDLNSLRRLSVHLELCTRSTTGTFD CNRFLDGVDKAPAVIKPAAAGQSSSLCDKV" 2137. .2568 /note="ORF5; Ac4/Op8" /codon_start=1 /product="unknown" /protein_id="AAP29799.1" /db_xref="GI:30269983" /translation="MPNEVKEFSLRLTRESKENIVAHIGHLSRVRELIDRNVTPADVRR FRFPDRLAALAAWVNVQAYGADGTIRLOPTLIYRVCRNFRADVAPADDFHFIAR YLMEPCGTPLVIDHPLHVFGEETEGVNELLEVRINAGDDL" 2669. .2860 /note="ORF6; Ac2" /codon_start=1 /product="baculovirus repeated ORF" /protein_id="AAP29800.1" /db_xref="GI:30269984" /translation="MDRNKKNAVTDTWIFKCIHYTRTGIHHPCLSNFVKRPVYLO PHTVLINNYAQDAPQFAI" complement(2896. .3699) /note="ORF7; Ac136/Op132" /codon_start=1 /product="p26" /protein_id="AAP29801.1" /db_xref="GI:30269985" /translation="MAMLTFLICVSSAALSVMKVTINNVLVTYNDTKTIAVKQVD DKPAIQVIPPQSFSTKQNEELMLHFFPGVASNVMPRIANNKLTLLSDGLATIT VDRVTNPHSHKRMVYQOLYSFALSFPANQIYIGAPIFEKRMVSVITARHEDYK NKLVIYPTVGTISARGLSVGINFDLQILTOKLEGSSVYGMQLPVKALKDYAISTNR KKNLFKGLPRNVAVFYNERDITIALVEGEFEIDRLRLSGPLILRNKQ"
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	complement(3738. .4220) /note="ORF8; ptp2; Op9" /codon_start=1 /product="protein tyrosine phosphatase 2" /protein_id="AAP29802.1" /db_xref="GI:30269986" /translation="MIDANQIDKVIYFGVGYGDDKAMLOFIKKYDIASVISLINADV GPIRQALGPAGEHIHYDCEDAPTVALPNAMFALYEYTRIGRIGKRVLIHCYAGESR SAALVYVYMRSRQMSYBEALSLVKNKRRVAISNHFVRFLASKCSKYFVNVLKIRVS " complement(4198. .4731) /note="ORF9; ptp1; Ac1/Op10" /codon_start=1 /product="protein tyrosine phosphatase 1" /protein_id="AAP29803.1" /db_xref="GI:30269987" /translation="MFPDRWHEFTPCGRVIDGTRLICFKVPLSASLEFVTTNDEDRWT VASLLTRHSALGAVIDLNTTRYDYDGEQMIREGLLYKKIRVPGRAIPEDTVOKEFSA VDEFQDRCTPLVGVGTHGLNSGILVCRIMVDKLGVSADATIRFEARGHKIERA NYLQDLARNHVRREP" 4800. .5819 /note="ORF10; Ac11/Op11" /codon_start=1 /product="unknown" /protein_id="AAP29804.1"
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE	complement(5819. .6420) /note="ORF11; Ac12/Op12" /codon_start=1 /product="unknown" /protein_id="AAP29805.1"

CDS

CDS

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	81.2	14.4	679	1	AV398660	AV398660 AV398660
2	39.6	7.0	954	4	CL077394	CL077394 CH216-145
C 3	39	6.9	762	9	EX147578	EX147578 Danilo rer
4	39	6.9	1324	9	CG751412	CG751412 P045-4-E0
C 5	38.6	6.8	758	8	BH517933	BH517933 BOGBL87TF
6	38.2	6.8	535	7	N63392	N63392 Y-235h10.81
C 7	38	6.7	578	8	BZ902838	BZ902838 CH240_23B
8	38	6.7	582	8	BZ899636	BZ899636 CH240_16P
9	37.2	6.6	425	6	CB410541	CB410541 NISC_nc10
10	37.2	6.6	669	7	CG384338	CG384338 AGENCOURT
C 11	36.8	6.5	1101	9	CNS0039X	AL063938 Drosophill
12	36.6	6.5	432	1	A1437474	A1437474 fb34b06.x
C 13	36.6	6.5	580	4	BI8433287	BI8433287 ft59f05.x
14	36.6	6.5	1048	3	CR709817	CR709817 Tetraodon
15	36.4	6.5	480	6	CB484581	CB484581 cclwtbDHO
16	36.4	6.5	857	9	CG935894	CG935894 MBEAV05TR
17	35.8	6.3	461	8	BH757407	BH757407 SALK_0561
18	35.8	6.3	655	9	AG358704	AG358704 Mus_muscu
19	35.8	6.3	675	4	BG595019	BG595019 NISC_iv11
20	35.8	6.3	715	2	BE374425	BE374425 501272596
21	35.8	6.3	1049	9	CL510716	CL510716 SAIL_8336
C 22	35.6	6.3	516	4	BM278655	BM278655 As_Egz_65
23	35.6	6.3	613	9	CL753134	CL753134 OR_BBa012
24	35.6	6.3	626	9	AG401197	AG401197 Mus_muscu

```

Query Match      14.4%; Score 81.2; DB 1; Length 679;
Best Local Similarity 71.3%; Pred. No. 2.1e-14;
Matches 107; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY      129  AAAAAACACGCTTTTGCACGGGGCCCATACATAGTACAAACTCTACGTTTCGTAGACTA 188
      |||||
Db       1  AAAAAAGTTCGCCTATCTTTTGACATATAATATACAGTACGAACTCTCAAAATCGTAGACTA 60
      |||||

```

```

Qy 189 TTTTACATAAAGTCTACACGGTTGTATACGGTCCAAATACACTACACACATTGAACC 248
Db 61 TTTTATTAGAATAGTCTACATGTACTACTACGGTCTCAATATACTACTACTATCAACT 120
Qy 249 TTTTTCAGTGCAGAAAAGTAGTCGTGCGC 278
Db 121 TTTTTCGATTACAAAAAGTTTCATTTTGC 150

RESULT 2
LOCUS CL077394
DEFINITION CH216-145B11_Sp6.1 CH216 Xenopus tropicalis genomic clone
ACCESSION CH216-145B11, genomic survey sequence.
VERSION CL077394
KEYWORDS GSS.
SOURCE CL077394.1 GI:40533307
ORGANISM Xenopus tropicalis (western clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 954)
AUTHORS Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTATAGTGACATATAG
Class: BAC ends
High quality sequence start: 163
High quality sequence stop: 226.
FEATURES
Location/Qualifiers
source 1..954
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-145B11"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/notes="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

ORIGIN
Query Match 7.0%; Score 39.6; DB 9; Length 954;
Best Local Similarity 52.4%; Pred. No. 0.48;
Matches 87; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 103 CCCACCACCACTTTTTCGACTCGAAACACGCTTTTGCAGCGGGCCCATACATA 162
Db 330 CGCCCCCCCCCTTTTTCAGAAAAACACCCCTCTCTACCCACCCCTTAAACATA 389
Qy 163 GTACAAACTCTACGTTTCGTAGACTATTTTACATAAATAGTCTACACGGTTGTATACGCT 222
Db 390 ATCTGAACCTCTATATCACTCATCTATTATATACGATAATGTACCCCTACTGTAAATC 449
Qy 223 CCAATACACTACACACATTTGAACCTTTTTCGAGTGCAGAAAAAGT 268
Db 450 ATAACGATATTACCACCTCACTGACTCTCTCTGTACCCATATAAAT 495

RESULT 3
LOCUS BX147578/c
DEFINITION Danio rerio genomic clone DKEY-109J14, genomic survey sequence.
ACCESSION BX147578

```

```

VERSION BX147578.1 GI:27978953
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 762)
AUTHORS Humphray, S.J., Huckle, E. and Durham, J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 109J14. 109J14
is part of the Daniokey BAC Library created by R. Piasterik and N.V.
Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES
Location/Qualifiers
source 1..762
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-109J14"
/tissue type="Testis"
/notes="vector pIndigoBAC-536"

ORIGIN
Query Match 6.9%; Score 39; DB 9; Length 762;
Best Local Similarity 56.7%; Pred. No. 0.71;
Matches 72; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 157 TACATAGTACAACTCTACGTTTCGTAGACTATTTTACATAAATAGTCTACACGGTTCTA 216
Db 655 TATAAATAACACAAAATCCTATTAGTCGACATTTTATAATCGTTTATAGCAAAAC 596
Qy 217 TACGCTCCAAATACATACACACATTTGCACTTGAACCTTTTGCAGTGCAGAAAAAGTACGTGTCG 276
Db 595 TCATTTCCAAAACACTACTAGTCAATTTGACGTCAGGAAATGCAATTCGTTAATATGTCA 536
Qy 277 GCAGTCA 283
Db 535 GTCGACA 529

RESULT 4
LOCUS CG751412
DEFINITION P045-4-E06.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
ACCESSION CG751412
VERSION CG751412.1 GI:37973841
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1324)
AUTHORS Srinivasan, J., Sinz, W., Jesse, T., Wiggers-Perebolte, L., Jansen, K.,
Buntjer, J., van der Meulen, M. and Sommer, R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus
pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES
Location/Qualifiers

```



```
aps 0;
TTGT 215
```

```

Db      218 ATGTATACCTTAACAGTTACCTTGTAATAAATAATCAACAAAAGACAAAACCAATTT 277
Qy      216 ATAGCTCCAAATACACTACACACATGAACCTTTTGGAGTGCAGAAAAGTACGTGTC 275
Db      278 ATTTGCCCAATTTACACAGGTTTTTACTTTAAATAATCTACCTTATGGGTAAATACCAAGC 337
Qy      276 GGCAGTCAGTAGCGCGGCTTATCGGTCGCGTCTGTGACGTAACGAATCACATTATCG 335
Db      338 TCCAATCGGTAAACCAACCTTGGCAATGTTTTTACCTACCACCTGGCCACCAAGGTTTTAG 397
Qy      336 GACCGGACGAGTGTGT 352
Db      398 GACNGGATAAATNATTT 414

RESULT 7
BZ902838/c
LOCUS      BZ902838
DEFINITION CH240_23B3.TJ CHORI-240 Bos taurus genomic clone CH240_23B3,
            genomic survey sequence.
ACCESSION  BZ902838
VERSION    BZ902838.1 GI:31627927
KEYWORDS   GSS.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE  1 (bases 1 to 578)
AUTHORS   Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,
            Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,
            Womack,J.E., de Jong,P.J. and Lewin,H.A.
TITLE     A Cattle-Human Comparative Map Built with Cattle BAC-ends and Human
            Genome Sequence
JOURNAL   Unpublished (2003)
COMMENT   Other GSSs: CH240_23B3.TV
            Contact: Harris Lewin
            Department of Animal Sciences
            University of Illinois at Urbana Champaign
            1201 W. Gregory Dr., Urbana, IL 61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.html). For BAC library
            availability, please contact Pieter de Jong (pjejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/orderinginformation.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by the University of Illinois at Urbana
            Champaign, USA with funds provided by grant No. AG202-34480-11828
            from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
            Genome Sequencing Initiative)
            Plate: 23 row: B column: 3
            Seq primer: SP6
            Class: BAC ends.
FEATURES   Location/Qualifiers
            source
              1..578
               /organism="Bos taurus"
               /mol_type="genomic DNA"
               /strain="breed: Hereford"
               /db_xref="taxon:9913"
               /clone="CH240_23B3"
               /sex="Male"
               /cell_type="Blood"
               /clone_lib="CHORI-240"
               /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
               Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
               library (Male) produced by Pieter de Jong"
ORIGIN
Query Match      6.7%; Score 38; DB 8; Length 578;
Best Local Similarity 55.2%; Pred. No. 1.4;

```

```

Best Local Similarity 55.2%; Pred. No. 1.4;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy      1 CATGATGATAAACAATGATATGGTGCTTAATGTGTTTCAACAACAATTTCTGTGAACGTGTG 60
Db      252 CATGCTGATCAATGAGGCATAAAGATATAATGTTTATAGACACATTTGTTGGAATGTA 193
Qy      61 TTTTCATGTTTGCACAACGACCTTTATACTCGGTGGCTCCCCACCACTTTT 120
Db      192 TTACCAGCTTTTAAACACCACATTTATGTTTGGGCCTAAATCATCAGCAATGAAT 133
Qy      121 GCACTGCACAAAAA 134
Db      132 TCAGCATTAATAA 119

RESULT 8
BZ899636
LOCUS      BZ899636
DEFINITION CH240_16P5.TV CHORI-240 Bos taurus genomic clone CH240_16P5,
            genomic survey sequence.
ACCESSION  BZ899636
VERSION    BZ899636.1 GI:31624687
KEYWORDS   GSS.
SOURCE     Bos taurus (cow)
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.,
            Bachman,S., Green,S., Campos,E.J., Benson,L.D., Edwards,J., Liu,L.,
            Womack,J.E., de Jong,P.J. and Lewin,H.A.
TITLE     A Cattle-Human Comparative Map Built with Cattle BAC-ends and Human
            Genome Sequence
JOURNAL   Unpublished (2003)
COMMENT   Contact: Harris Lewin
            Department of Animal Sciences
            University of Illinois at Urbana Champaign
            1201 W. Gregory Dr., Urbana, IL 61801, USA
            Tel: 217 333 5998
            Fax: 217 244 5617
            Email: h-lewin@uiuc.edu
            Clones are derived from the bovine BAC library CHORI-240
            (http://www.chori.org/bacpac/bovine240.html). For BAC library
            availability, please contact Pieter de Jong (pjejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/orderinginformation.htm). This work
            was undertaken as part of the International Bovine BAC Mapping
            Consortium (IBBMC) by the University of Illinois at Urbana
            Champaign, USA with funds provided by grant No. AG202-34480-11828
            from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
            Genome Sequencing Initiative)
            Plate: 16 row: P column: 5
            Seq primer: T7
            Class: BAC ends.
FEATURES   Location/Qualifiers
            source
              1..582
               /organism="Bos taurus"
               /mol_type="genomic DNA"
               /strain="breed: Hereford"
               /db_xref="taxon:9913"
               /clone="CH240_16P5"
               /sex="Male"
               /cell_type="Blood"
               /clone_lib="CHORI-240"
               /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
               Hereford bull L1 Domino 99375; CHORI-240 Bovine BAC
               library (Male) produced by Pieter de Jong"
ORIGIN
Query Match      6.7%; Score 38; DB 8; Length 582;
Best Local Similarity 55.2%; Pred. No. 1.4;

```

```

Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 CATGATGATAAACAATGTATGGTCTTAATGTTGCTTCAACAACAATCTGTGAACGTGG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 CATGCTGATCAATGAGGCATAAGATATAATTTGTTTATAGACACATTTGTTGGAATGTA 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 TTTTTCATGTTTCCACACAGCACCTTTATCTACGTGGTGGCTCCACACCAACTTTTT 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 TTACCAGCTTTTAACACACCACATTTATGTTTGGGCGCTAAATCATCAGCAAAATGAAT 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 GCACGCTCAAAAAA 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 TCAGCATTTAAAAA 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
LOCUS CB410241
DEFINITION NISC_nc10d09.x1 COGENE 6E MAX Homo sapiens cDNA clone IMAGE:5776553
3', mRNA sequence.
ACCESSION CB410241
VERSION CB410241.1 GI:29166981
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 425)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM12849 row: G column: 18
Seq primer: -21M13 forward primer (ABI).
FEATURES
Location/Qualifiers
1..425
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5776553"
/tissue_type="maxilla, pooled"
/dev_stage="6 weeks postconception"
/lab_host="DH10B"
/clone_lib="COGENE 6E MAX"
/note="Vector: pAMP1; cDNA primed using oligo-dT primer,
directionally cloned into UDG sites of pAMP1. Size
selected for insert sizes ranging from 0.2-1.8 kb.
Normalized to Cot5. Primary library, non-amplified.
Library constructed by M. Lovett. For more information on
this library, please contact R. Tidwell (Washington
University) or visit the COGENE website at
http://hg.wustl.edu/COGENE/."

ORIGIN
Query Match 6.6%; Score 37.2; DB 6; Length 425;
Best Local Similarity 54.3%; Pred. No. 2;
Matches 75; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 115 TTTTTCGACTCCAAAAACACGCTTTTGCACGCGGCCCATACATAGTACAACTCTA 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 TTTTTCGACTCCAAAAACGCTTTTTCCTCGGGTTTTATCCACTGTCAATACTGTA 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 175 CTTTTCGCTAGACTATTTACATAAATAGTCTACACGGTTGTATACGCTCCCAATACACTA 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 126 TTTTGTGCGAATATATTTCGCAAAAAAACTCAGCTTTTATTTCCATTTAAACAACCTA 185
QY 235 CCACACATTGAACCTTTT 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 CAATATTACAAGCTGTT 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
LOCUS CO384338
DEFINITION laevis cDNA clone IMAGE:7297357 3', mRNA sequence.
ACCESSION CO384338
VERSION CO384338.1 GI:49490161
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE 1 (bases 1 to 669)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-remail.nih.gov
Tissue Procurement: Bruce Blumberg
cDNA Library Preparation: B. Blumberg
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM15322 row: j column: 11
High quality sequence stop: 534.
Location/Qualifiers
1..669
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:7297357"
/sex="both"
/tissue_type="dorsal blastopore lip"
/lab_host="TOP10"
/clone_lib="Blumberg, Cho dorsal blastopore lip"
/note="Organ: embryo; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Library was prepared from 50 ug of
total RNA by oligo-dT priming and AMV reverse
transcriptase. After addition of EcoRI linkers and
EcoRI-XhoI digestion, the cDNA was size selected by
chromatography on Sepharose CL-4B columns and fractions
containing cDNAs larger than 500 bp were ligated into
EcoRI-XhoI-digested lambda ZAPII (UnizAP-XR) and packaged
in vitro. Average insert size is 1.4 kb. The original
library contained 6 x 106 recombinants, of which 3 x 106
were amplified and stored at -70 C in SM buffer containing
7% DMSO. 3 x 106 pfu were mass excised and the resulting
phagemids used to infect Top10F. References: Science 253,
196-196 and Methods in Molecular Biology 97, 555-574.
Additional sequences from this library have been deposited
under the name Xenopus laevis dorsal blastopore lip.
Library constructed by Bruce Blumberg (University of
California, Irvine, Department of Developmental and Cell
Biology)."

ORIGIN
Query Match 6.6%; Score 37.2; DB 7; Length 669;
Best Local Similarity 57.9%; Pred. No. 2.6;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```

```

QY 153 CCCATACAGTACAACTCTACGTTGGTAGACTATTTTACATAAATAGTCTACACGGT 212
Db 104 CCATAAATAGGGAACCTCTCGTATATAGTTGTAAGAAATACATCCATGATGCAACGT 163

QY 213 TGTATAGCTCCAATACACTACACATCCACATTTGAGACCTTTTTCGAGTGCAGAAAAA 266
Db 164 TGTAGTCTCAGAAATACATTTTAATCTTTTTCCTTTTTCGAGTGCAGAAATA 217

RESULT 11
CNS0039X/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BACR08012 of RPCL-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063938
VERSION AL063938.1 GI:4941795
SOURCE GSS.
ORGANISM Drosophila melanogaster (fruit fly)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pletier de Jong's laboratory in the Department
of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCL-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08012"
/clone_lib="RPCL-98"
/note="end : T7"

ORIGIN
Query Match 6.5%; Score 36.8; DB 9; Length 1101;
Best Local Similarity 35.9%; Pred. No. 3.9; Mismatches 74; Indels 0; Gaps 0;
Matches 61; Conservative 35;

QY 73 CCAACAAGCACCTTTATCTCGGTGGCGCTCCCAACCACTTTTTCGACTGCAGAAAA 132
Db 1068 CATTMMATMACACATATATCTCTCTMTCTMYACAMACAAATWACTAYWCACCTMTAYA 1009

QY 133 AACACGCTTTTGCACGGGGCCCATACATAGTACAACTCTACGTTTCGTAGACTATTTT 192
Db 1008 CMCACYAYWCACAAACMCCTCCCTCTCTAATAAACATCTAATAATCAWAAATACMAH 949

QY 193 ACATAAATAGTCTACACCGCTTGTATACGCTCCAAATACACTACACACAT 242
Db 948 MAMAATAWATACWAWYHWTTCACACACWCACTATACWCAWACYACWT 899

RESULT 12
AT437474

```

```

LOCUS
DEFINITION AT437474 432 bp mRNA linear EST 07-JUN-2001
fb34b06.x1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone
IMAGE:3713747.3, mRNA sequence.
ACCESSION AT437474
VERSION AT437474.1 GI:4286113
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 432)
AUTHORS Clark, M., Johnson, S. L., Lehtach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
CNSA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Ressourcenzentrum PrimatDatenbank, Berlin, Germany (web address:
www.rzpdb.de)
Seq primer: T7 ET from Amersham
High quality sequence stop: 423
POLYA=No.
FEATURES
Location/Qualifiers
1..432
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:3713747"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/clone_lib="Zebrafish WashU MPIMG EST"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; 1st
strand cDNA was primed with a Not I - oligo (dT)15 primer
[5'pGACTAGTTCTAGATCGGAGCGCGCCCTTTTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
ORIGIN
Query Match 6.5%; Score 36.6; DB 1; Length 432;
Best Local Similarity 50.9%; Pred. No. 3.7; Mismatches 84; Indels 0; Gaps 0;
Matches 87; Conservative 0;

QY 109 ACCAACTTTTTCGCACTGCAAAAAAACACGCTTTTTCACGCGGGCCCATACATAGTACAA 168

```

	Matches	87;	Conservative	0;	Mismatches	84;	Indels	0;	Gaps	0;
QY	109	ACCAACTTTTTTGGCACTGCAAAAAAACACGCTTTTTCACGCGGGCCATACATAGTACAA	168							
Db	27	ATCACTTATTTCTCTACTGTGTACACGCGTAATTTTACTTTTGCAAGTGACAAAACTG	86							
QY	169	ACTCTACGTTTCGTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAAT	228							
Db	87	AGTGAACAATTTACAAAATATATAAAACAACAACGTCTTCTGTGCTACTAAGCACTT	146							
QY	229	ACACTACCAACAATTGAACCTTTTTCAGTGTCAAAAAAGTACGTGTGGCA	279							
Db	147	ACTCTACTACACACTTCAATTACACGGAGAGCGCTAAAAGTCATATGACAGAA	197							

RESULT 14			
CR709617/c			
LOCUS	CR709617	1048 bp	linear
DEFINITION	Tetraodon nigroviridis full-length cDNA.		
			HTC 19-AUG-2004

ACCESSION	CR709617
VERSION	CR709617.1
KEYWORDS	HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE	Tetraodon nigroviridis
ORGANISM	Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.

1 (bases 1 to 1048)

REFERENCE

REFERENCE	Retraquembourg, A. Retraquembourg, A. Retraquembourg.
AUTHORS	1 (bases 1 to 1048)
TITLE	Genoscope.
JOURNAL	Direct Submission Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage CA 5206 - 01057 Evry cedex FRANCE 2 rue Gaston Cremieux

COMMENT

(E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
2 rue Gaston Crémieux, CS306 2 JAVOY EVRY CEDEX - FRANCE

The sequences are based on single pass reads.
More information available at
<http://www.genoscope.cns.fr/tetraodon>.

```
FEATURES
source
Location/Qualifiers
1..1048
/organism="Tetraodon nigroviridis"
/mol_type="mRNA"
```

ORIGIN
/db_xref="taxon:99883" /tissue_type="Eggs"

Query Match 6.5%; Score 36.6; DB 3; Length 1048;
Best Local Similarity 49.5%; Pred. No. 4.5;
Matches 93; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 22 GTGCTAATGTCCTTCAACAATACTGTGAACGTGTTTTCATGTTTGGCCACAAGC 81
 Db 1033 GTGGAAAGATGCTTTTNAATTACACGCCAATTCACATGCTCTTGGTTTGGACAGACTC 974

Qy 82 ACCTTTATACTCGGTGGCTCCCAACCAACTTTTTTGCACTCGAAAAAACAGCTT 141
| | | | |
Db 973 GACCATCGACTGGGTCCAAATCATCTCAGGGCAAGAAAATCCATTATTACAGCA 914
| | | | |

QY
Db

142 TTGCACGGCGGCCCTACATAGTACAACCTCAGTTTCGTGACACTATTTTACATAAATA 201
913 GTTAATGGCACTAACACCCTAATTCAAANAATTGAGTTTATTAGTTTTTTTAA'TTATAAA 854

Qy	202	GTCTACAC	209
Db	853	CTCTTCAC	846

DD FORM 855 JUL 1963

RESULT 15
CB484581

CB484581	CB484581	480 bp	mRNA	linear	EST 01-APR-2003
LOCUS	cclwtb09074				
DEFINITION	Coregonus clupeaformis head				
	CDNA, mRNA sequence.				
ACCSSION	CB484581				

ACCESSION	CB484581
VERSION	CB484581.1
	GI:29295807

FEATURES	source	Location/Qualifiers
		1..480
		/organism="Coregonus clupeaformis"
		/mol_type="mRNA"
		/db_xref="taxon:59861"
		/clone_lib="Coregonus clupeaformis head"
		/note="Vector: pBluescriptIISK+; Library Creator: Kristian R von Schalburg ; Lake whitefish tissue contributor: Louis Bernatchez (Universite Laval)"
ORIGIN		
	Query Match	6.5%; Score 36.4; DB 6; Length 480;
	Best Local Similarity	51.9%; Pred. No. 4.4;
	Matches	82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
OY	49	TGTTGAACGTGTGTTTTCATGTTTGCCAAACGAAGCACCTTTTATATCGGTGGCTCCCCACC 108
Db	10	TTTGTGAATGTGCTTTTATTTGTTTCCCAATCATCCAACGTTTCTATCATCAATCCCAACATCCA 69
OY	109	ACCAACTTTTTTTCACCTGTCGAAAAAACAACGCTTTTTCGACGCGGGCCCATACATAGTACAA 168
Db	70	GACAAATTTTATCGTAAATATTAAGTCACAAATTCAGCAACTGTATGCTTACTAAATACAA 129
OY	169	ACTCTACGTTTCGTAGACTATTTTACATAAATAGTCTA 206
Db	130	CACCTGGGTATATCGAAGCCATTTTACAGAAATGGACTA 167

Search completed: October 24, 2005, 21:53:16
Job time : 4081.35 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:39:18 ; Search time 616.051 Seconds
(without alignments)
5419.578 Million cell updates/sec

Title: US-09-896-888A-1

Perfect score: 564

Sequence: 1 catgtatataacaatgat.....tggtacagcgacacaacatg 564

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2002bs.*

8: Geneseqn2003as.*

9: Geneseqn2003bs.*

10: Geneseqn2003cs.*

11: Geneseqn2003ds.*

12: Geneseqn2004as.*

13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	564	100.0	564	2 AAV62487	AAV62487 O. pseudo
2	548	97.2	560	12 ADQ48575	Adq48575 Opiet2 pro
3	548	97.2	2773	9 AAL61306	Aal61306 p22Op2F e
4	546.4	96.9	5038	12 ADQ48539	Adq48539 Viral vec
5	101.4	18.0	279	2 AAT13730	Aat13730 AcNPV ORF
6	101.4	18.0	133894	2 AAT13635	Aat13635 AcNPV gen
7	93.4	16.6	28413	10 ADC51646	Adc51646 BmNPV gen
8	62	11.0	141	12 ADQ48576	Adq48576 Viral vec
9	39.6	7.0	6289	8 ABZ10059	Abz10059 Haematopo
10	39.6	7.0	9289	4 AAS46501	Aas46501 Tumour su
11	39.6	7.0	9289	10 ADE84121	Ade84121 Human lym
12	38	6.7	9289	8 ABZ10205	Abz10205 Haematopo
13	38	6.7	9289	10 ADE84197	Ade84197 Human lym
14	36.2	6.4	2000	8 ADZ71938	Adz71938 Rice gene
15	35.4	6.3	986	6 ABQ68715	Abq68715 Listeria
16	35.4	6.3	1549	6 ABQ70339	Abq70339 Listeria
17	34.4	6.1	855	2 AAV34232	Aav34232 Human sec
18	34.4	6.1	855	8 ACD08103	Acd08103 cDNA enco
19	34	6.0	18977	4 ABL19380	Ab119380 Drosophil
20	33.8	6.0	1991	6 AAS18541	Aas18541 Partial g

ALIGNMENTS

RESULT 1

AAV62487

ID AAV62487 standard; DNA; 564 BP.

XX AAV62487;

AC AAV62487;

DT 17-OCT-2003 (revised)

DT 19-JAN-1999 (first entry)

XX

DE O. pseudotsugata multcapsid nucleopolyhedrosis virus ie2 promoter.

XX

KW Orgyia pseudotsugata; multcapsid; nucleopolyhedrosis virus; OpNPV;

KW Op ie2; promoter; shuttle vector; transformation; melanotransferrin;

KW immediate early baculovirus promoter; prokaryotic; transcription;

KW bleomycin/phleomycin-type antibiotic; insect cell; transposon;

KW ion transport peptide hormone; ss.

XX

OS Orgyia pseudotsugata; polyhedrosis virus.

XX

PN WO9844141-A2.

XX

PD 08-OCT-1998.

XX

PF 26-MAR-1998; 98WO-CA000282.

XX

PR 27-MAR-1997; 97US-0049946P.

PR

PR 28-JAN-1998; 98CA-02221819.

XX

PA (UYBR-) UNIV BRITISH COLUMBIA.

XX

PI Grigliatti TA, Theilmann DA, Pfeifer TA, Hegedus DD;

XX

XX WPI; 1998-557129/47.

XX

XX Expression vectors for transforming insect cells from disparate lines -

XX useful to express heterologous DNA, e.g. to allow study of gene

XX expression and produce commercially important proteins.

PS

PS Claim 10; Page 82; 121bp; English.

CC

CC This represents a Orgyia pseudotsugata multcapsid nucleopolyhedrosis

CC virus (OpNPV) immediate early 2 (Op ie2) promoter sequence. The

CC invention provides a new shuttle vector for transforming insect cells

CC that comprises: (i) prokaryotic origin of replication; (ii) insect
CC promoter having homology to, and capable of functioning as, an immediate
CC early baculovirus promoter; (iii) prokaryotic promoter sequence, and (iv)
CC selectable marker capable of conferring resistance to a bleomycin/
CC phleomycin-type antibiotic under transcriptional control of (ii) and
CC (iii), in insect and prokaryotic cells respectively. The vectors can be
CC used to stably transform (especially insect) cells with heterologous DNA,
CC useful to allow study of gene expression and direct expression of
CC heterologous gene products e.g. commercially important proteins. They are
CC especially useful to allow expression of the heterologous
CC melanotransferrins, ion transport peptide hormones or biologically active
CC derivatives in insect cells. They enable transformation of insect cell
CC lines from disparate species, allowing screening of lines for optimum
CC post-translational modification of particular proteins. Shuttle vectors
CC further comprising DNA transposable elements defining a transposon can be
CC used to optimise heterologous protein expression and facilitate selection
CC of desired transformants. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
SQ Sequence 564 BP; 147 A; 160 C; 116 G; 141 T; 0 U; 0 Other;

Query Match 100.0%; Score 564; DB 2; Length 564;
Best Local Similarity 100.0%; Pred. No. 1.1e-172;
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGATGATACAAATGATGGTCTAATGTTGCTTCAACAACAATCTGTGAACCTGTG 60
Db 1 CATGATGATACAAATGATGGTCTAATGTTGCTTCAACAACAATCTGTGAACCTGTG 60
QY 61 TTTTCATGTTTGGCCAAAGCAGCCTTTTATCTCGTGGCCCTCCCAACCACTTTT 120
Db 61 TTTTCATGTTTGGCCAAAGCAGCCTTTTATCTCGTGGCCCTCCCAACCACTTTT 120
QY 121 GCATGCAAAAAAACAACGCTTTTGACCGCGGCCATACATAGTACAAACTCTACGTTTC 180
Db 121 GCATGCAAAAAAACAACGCTTTTGACCGCGGCCATACATAGTACAAACTCTACGTTTC 180
QY 181 GTAGACTATTTTTACATAAATAGTCTACACCGTCTGTATACGCTCCCAATACACTACACAC 240
Db 181 GTAGACTATTTTTACATAAATAGTCTACACCGTCTGTATACGCTCCCAATACACTACACAC 240
QY 241 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGGAGTCACGTAGCGCGGCTTATC 300
Db 241 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGGAGTCACGTAGCGCGGCTTATC 300
QY 301 GGGTGGCTCTGTACGATCAATTCATATTCGACCGGACGAGTGTCTTATCGT 360
Db 301 GGGTGGCTCTGTACGATCAATTCATATTCGACCGGACGAGTGTCTTATCGT 360
QY 361 GACAGGACCGCAGCTTCTGTGTTGCTAACCAGCGGACGCAACTCTTATCGGAACA 420
Db 361 GACAGGACCGCAGCTTCTGTGTTGCTAACCAGCGGACGCAACTCTTATCGGAACA 420
QY 421 GGACGCGCTCCATATACGCGCGGTTTATCTCATGCGCGTGAACGACGAGCGGCC 480
Db 421 GGACGCGCTCCATATACGCGCGGTTTATCTCATGCGCGTGAACGACGAGCGGCC 480
QY 481 GTCCGCTTATCGGCTTATAAATACAGCCCGGACGATCTGGTAAACACAGTTGAACAG 540
Db 481 GTCCGCTTATCGGCTTATAAATACAGCCCGGACGATCTGGTAAACACAGTTGAACAG 540
QY 541 CATCTGTTTACAGCGACACAACATG 564
Db 541 CATCTGTTTACAGCGACACAACATG 564

RESULT 2
ADQ48575
ID ADQ48575 standard; DNA; 560 BP.
XX
AC ADQ48575;
XX
DT 09-SEP-2004 (first entry)

XX OpIE2 promoter DNA sequence.
DE
XX
XX viral vector; recombination site; recombinant virus;
KW replication-defective particle generation; gene expression inhibition;
KW gene therapy vector; ds; OpIE2 promoter.
XX
XX Unidentified.
XX
XX WO2004009768-A2.
PN
XX 29-JAN-2004.
XX
XX 18-JUL-2003; 2003WO-US022437.
XX
XX 18-JUL-2002; 2002US-0396335P.
PR 26-JUL-2002; 2002US-0398617P.
PR 19-NOV-2002; 2002US-0427231P.
PR 24-MAR-2003; 2003US-0456496P.
PR 03-JUN-2003; 2003US-0474940P.
XX
XX (INVI-) INVITROGEN CORP.
PA (BENNI-) BENNETT R P.
PA (WELC/) WELCH P J.
PA (HARW/) HARWOOD S.
PA (MADD/) MADDEN K.
PA (FRIM/) FRIMPONG K.
PA (FRAN/) FRANKIE K E.
XX
XX Bennett RP, Welch PJ, Harwood S, Madden K, Frimpong K, Franke KE;
PI WPI; 2004-132944/13.
XX
XX Novel nucleic acid molecule comprising all or a portion of one or more
PT viral genome and further comprising two or more recombination sites that
PT do not substantially recombine with each other, useful as gene therapy.
XX
XX Disclosure; Fig 16; 555pp; English.
XX
XX The invention comprises a nucleic acid molecule consisting of all or a
CC portion of one or more viral/baculoviral genomes and further containing
CC at least two recombination sites that do not substantially recombine with
CC each other. One or more of the recombination sites is capable of
CC undergoing recombination with a compatible recombination site in the
CC presence of one or more proteins active in lambda recombination. The
CC nucleic acid of the invention replicates in prokaryotic and eukaryotic
CC cells. The nucleic acid of the invention is useful for constructing a
CC recombinant virus, generating replication-defective particles,
CC preventing/inhibiting the expression of one or more genes in an organism,
CC and are useful as gene therapy vectors. The nucleic acid of the invention
CC is also useful for producing and expressing fusion polypeptides. The
CC present DNA sequence represents the OpIE2 promoter that was used in the
CC exemplification of the invention.
XX
SQ Sequence 560 BP; 144 A; 156 C; 116 G; 144 T; 0 U; 0 Other;

Query Match 97.2%; Score 548; DB 12; Length 560;
Best Local Similarity 100.0%; Pred. No. 1.8e-167;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGATCATAAACAATGATGGTGTCTAATGTTGCTTCAACAACAATCTGTGAACCTGTG 60
Db 5 CATGATCATAAACAATGATGGTGTCTAATGTTGCTTCAACAACAATCTGTGAACCTGTG 64
QY 61 TTTTCATGTTTGGCCAAAGCAGCCTTTTATCTCGTGGCTCCCAACCACTTTT 120
Db 65 TTTTCATGTTTGGCCAAAGCAGCCTTTTATCTCGTGGCTCCCAACCACTTTT 124
QY 121 GCATGCAAAAAAACAACGCTTTTGACCGCGGCCATACATAGTACAAACTCTACGTTTC 180
Db 125 GCATGCAAAAAAACAACGCTTTTGACCGCGGCCATACATAGTACAAACTCTACGTTTC 184
QY 181 GTAGACTATTTTTACATAAATAGTCTACACCGTGTGTATACGCTCCCAATACACTACACAC 240

Db 185 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACACAC 244
Qy 241 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGTCGGCAGTCACGTAGCGCGCCTTATC 300
Db 245 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGTCGGCAGTCACGTAGCGCGCCTTATC 304
Qy 301 GGGTCGGTCTCTGTCAGTACGAATACATATCGGACCGGACGAGTGTGTCTTATCGT 360
Db 305 GGGTCGGTCTCTGTCAGTACGAATACATATCGGACCGGACGAGTGTGTCTTATCGT 364
Qy 361 GACAGCGCCAGCTTCTGTTGCTTAACCGCAGCGGACCACTCCTTATCGGAACA 420
Db 365 GACAGCGCCAGCTTCTGTTGCTTAACCGCAGCGGACCACTCCTTATCGGAACA 424
Qy 421 GGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGCGTGACCGGACACGAGCGGCC 480
Db 425 GGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGCGTGACCGGACACGAGCGGCC 484
Qy 481 GTCCCGCTTATCGCGCTATAAATACAGCCGCAACGATCTGGTAAACACAGTTGAACAG 540
Db 485 GTCCCGCTTATCGCGCTATAAATACAGCCGCAACGATCTGGTAAACACAGTTGAACAG 544
Qy 541 CATCTGTT 548
Db 545 CATCTGTT 552

RESULT 3

ADQ48539 standard; DNA; 2773 BP.
XX AAL61306;
XX AAL61306;
DT 22-SEP-2003 (first entry)
XX p220p2F expression vector for insect cells.
DE Multimeric protein; interleukin 5; IL5; TNFalpha; inflammatory disease;
KW tumour necrosis factor alpha; gene therapy; arthritis; gene; ds.
XX Unidentified.
OS WO2003042244-A2.
PN 16-NOV-2001; 2001US-0331575P.
XX 22-MAY-2003.
PF 15-NOV-2002; 2002WO-DK000764.
XX 16-NOV-2001; 2001DK-00001702.
PR 16-NOV-2001; 2001US-0331575P.
XX (PHAR-) PHARMEXA AS.
PA (KLYS/) KLYSNER S.
PA (NIEL/) NIELSEN F S.
PA (BRAT/) BRATT T.
PA (VOLD/) VOLDORGB B.
PA (MOUR/) MOURITSEN S.
XX Klyener S, Nielsen FS, Bratt T, Voldborg B, Mouritsen S;
XX WPI; 2003-449558/42.
DR New immunogenic analogue of a polymeric protein, useful for preparing a
PT composition for treating inflammatory diseases e.g. arthritis.
XX Disclosure; Page 195-196; 196pp; English.
XX The invention relates to immunogenic analogues of multimeric proteins
CC such as immunogenic variants of interleukin 5 (IL5) and tumour necrosis
CC factor alpha (TNF, TNFalpha) and methods for production of immunogenic
CC analogues. The immunogenic analogue is useful for preparing a composition
CC for treating inflammatory diseases, e.g., arthritis. It is also used in

CC gene therapy. The present sequence is p220p2F expression vector for
CC insect cells. This sequence is used to illustrate the method of the
CC invention
SQ Sequence 2773 BP; 669 A; 750 C; 673 G; 681 T; 0 U; 0 Other;
Query Match 97.2%; Score 548; DB 9; Length 2773;
Best Local Similarity 100.0%; Pred. No. 3.8e-167;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CATGATGATAAACAATGATGTCCTAATGTTGCTTCAACAACAATCTCTGTTGAACGTG 60
Db 5 CATGATGATAAACAATGATGTCCTAATGTTGCTTCAACAACAATCTCTGTTGAACGTG 64
Qy 61 TTTTCATGTTTGCACCAAGCACCTTTTATCTCGGTGCTCCGCCACCACTTTTTC 120
Db 65 TTTTCATGTTTGCACCAAGCACCTTTTATCTCGGTGCTCCGCCACCACTTTTTC 124
Qy 121 GCATGCAAAAAACACGCTTTTGACCGGGGCCATACATAGTACAACTCTAGCTTTC 180
Db 125 GCATGCAAAAAACACGCTTTTGACCGGGGCCATACATAGTACAACTCTAGCTTTC 184
Qy 181 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACACAC 240
Db 185 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACACAC 244
Qy 241 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGTCGGCAGTCACGTAGCGCGCCTTATC 300
Db 245 ATTGAACCTTTTTCAGTGCAAAAAAGTACGTGTCGGCAGTCACGTAGCGCGCCTTATC 304
Qy 301 GGGTCGGTCTCTGTCAGTACGAATACATATCGGACCGGACGAGTGTGTCTTATCGT 360
Db 305 GGGTCGGTCTCTGTCAGTACGAATACATATCGGACCGGACGAGTGTGTCTTATCGT 364
Qy 361 GACAGCGCCAGCTTCTGTTGCTTAACCGCAGCGGACCACTCCTTATCGGAACA 420
Db 365 GACAGCGCCAGCTTCTGTTGCTTAACCGCAGCGGACCACTCCTTATCGGAACA 424
Qy 421 GGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGCGTGACCGGACACGAGCGGCC 480
Db 425 GGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGCGTGACCGGACACGAGCGGCC 484
Qy 481 GTCCCGCTTATCGCGCTATAAATACAGCCGCAACGATCTGGTAAACACAGTTGAACAG 540
Db 485 GTCCCGCTTATCGCGCTATAAATACAGCCGCAACGATCTGGTAAACACAGTTGAACAG 544
Qy 541 CATCTGTT 548
Db 545 CATCTGTT 552
RESULT 4
ADQ48539
ID ADQ48539 standard; DNA; 5038 BP.
XX AC ADQ48539;
XX DT 09-SEP-2004 (first entry)
XX DE Viral vector-related plasmid - pIB/V5-His-DEST.
XX KW viral vector; recombination site; recombinant virus;
KW replication-defective particle generation; gene expression inhibition;
XX gene therapy vector; ds; plasmid.
OS Unidentified.
XX WO2004009768-A2.
PN 29-JAN-2004.
PD 18-JUL-2003; 2003WO-US022437.
XX

```
PR 18-JUL-2002; 2002US-0396335P.
PR 26-JUL-2002; 2002US-0398617P.
PR 19-NOV-2002; 2002US-0427231P.
PR 24-MAR-2003; 2003US-0456496P.
PR 03-JUN-2003; 2003US-0474940P.
XX (INVI-) INVITROGEN CORP.
PA (BENN-) BENNETT R P.
PA (WELC-) WELCH P J.
PA (HARW-) HARWOOD S.
PA (MADD-) MADDEN K.
PA (FRIM-) FRIMPPONG K.
PA (FRAN-) FRANK K E.
XX
PI Bennett RP, Welch PJ, Harwood S, Madden K, Frimpong K, Franke KE;
XX WPI; 2004-132944/13.
DR
XX
XX Novel nucleic acid molecule comprising all or a portion of one or more
PT viral genome and further comprising two or more recombination sites that
PT do not substantially recombine with each other, useful as gene therapy.
XX
XX Example 18; Page 395-403; 555pp; English.
XX
XX The invention comprises a nucleic acid molecule consisting of all or a
XX portion of one or more viral/baculoviral genomes and further containing
XX at least two recombination sites that do not substantially recombine with
XX each other. One or more of the recombination sites is capable of
XX undergoing recombination with a compatible recombination site in the
XX presence of one or more proteins active in lambda recombination. The
XX nucleic acid of the invention replicates in prokaryotic and eukaryotic
XX cells. The nucleic acid of the invention is useful for constructing a
XX recombinant virus, generating replication-defective particles,
XX preventing/inhibiting the expression of one or more genes in an organism,
XX and are useful as gene therapy vectors. The nucleic acid of the invention
XX is also useful for producing and expressing fusion polypeptides. The
XX present DNA sequence represents a plasmid that was used in the
XX exemplification of the invention.
XX
XX Sequence 5038 BP; 1342 A; 1178 C; 1198 G; 1320 T; 0 U; 0 Other;
SQ
```

```
Query Match          96.9%; Score 546.4; DB 12; Length 5038;
Best Local Similarity 99.8%; Pred. No. 1.7e-166;
Matches 547; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 CATGATGATAAACAATGATGGTCTAATGTTGTTCTTCAACAACAATTCGTTGAACCTGTG 60
DB 1 CATGATGATAAACAATGATGGTCTAATGTTGTTCTTCAACAACAATTCGTTGAACCTGTG 60
QY 61 TTTTCATGTTTGGCAACAGCACCTTTTATCTCGTGGCCCTCCACACCACTTTTTT 120
DB 61 TTTTCATGTTTGGCAACAGCACCTTTTATCTCGTGGCCCTCCACACCACTTTTTT 120
QY 121 GCACCTGCAAAAAAACAACACCTTTTGCACGCGGGCCCATACATAGTACAACTCTACGTTTC 180
DB 121 GCACCTGCAAAAAAACAACACCTTTTGCACGCGGGCCCATACATAGTACAACTCTACGTTTC 180
QY 181 CTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACCAAC 240
DB 181 CTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACCAAC 240
QY 241 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGTGGCAGTCAGTGGCGGCGCTTATC 300
DB 241 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGTGGCAGTCAGTGGCGGCGCTTATC 300
QY 301 GGGTGGCGTCTGTGTCAGTACGAATACATATTCGACCGGACGAGTGTGCTTTATCGT 360
DB 301 GGGTGGCGTCTGTGTCAGTACGAATACATATTCGACCGGACGAGTGTGCTTTATCGT 360
QY 361 GACAGGACCGCAGCTTCTGTGTTGCTTAACCGAGCGGACGCACTCCTTATCGGAACA 420
DB 361 GACAGGACCGCAGCTTCTGTGTTGCTTAACCGAGCGGACGCACTCCTTATCGGAACA 420
```

```
QY 421 GGACGCGCTCCATATCAGCCGCGGTATCTCATGCGGTGACCGACACAGGCGCCCC 480
DB 421 GGACGCGCTCCATATCAGCCGCGGTATCTCATGCGGTGACCGACACAGGCGCCCC 480
QY 481 GTCCCGCTTATCGCGCTATAAATACAGCCCGCAACGATCTGGTAAACACAGTTGAACAG 540
DB 481 GTCCCGCTTATCGCGCTATAAATACAGCCCGCAACGATCTGGTAAACACAGTTGAACAG 540
QY 541 CATCTGTT 548
DB 541 CATCTGTT 548
XX
XX RESULT 5
XX AAT13730/c
ID AAT13730 standard; DNA; 279 BP.
XX AC AAT13730;
XX
XX 16-OCT-2003 (revised)
DT 06-SEP-1996 (first entry)
XX
XX ACNPV ORF 152, residues 132387-132109.
XX
XX Autographa californica nuclear polyhedrosis virus clone 6; disruption;
XX non-essential gene; heterologous protein production; expression vector;
XX baculovirus; ss.
XX
XX Autographa californica nucleopolyhedrovirus; clone 6.
XX
XX Key Location/Qualifiers
CDS 1..279 /tag= a
FT /number= ORF 152
FT /note= "corresponds to AcNPV nucleotides complement
FT (132387-132109)"
XX
XX WO9601320-A2.
XX
XX 18-JAN-1996.
XX
XX 30-JUN-1995; 95WO-IB000578.
XX
XX 04-JUL-1994; 94GB-00013420.
XX
XX (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
XX
XX Bishop D, Possee R, Ayres M;
XX
XX WPI; 1996-087670/09.
XX
XX Autographa californica nuclear polyhedrosis virus complete genome
XX sequence - useful in the prodn. of vectors for enhanced heterologous
XX protein expression, such as interleukin(s), interferon(s) and
XX neurotoxin(s).
XX
XX Claim 1; Page 90-186; 122pp; English.
XX
XX AAT13636-731 show open reading frames 13, 20, 22-30, 32, 38, 41-46, 50-
XX 60, 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-127, 129-130, 140
XX - 146, 148-150, 152 and 154 from a total of 154 ORFs identified in the
XX Autographa californica nuclear polyhedrosis virus (AcNPV) clone 6. Each
XX gene is numbered according to its position in the virus genome beginning
XX at the left end of the linear map, and irrespective of its orientation.
XX The direction of transcription is relative to that of the polyhedrin
XX gene. Of the 154 ORFs identified it was found that some of the ORFs (ORFs
XX 27, 30, 32, 71, 86, 123, 126 and 127) are dispensable for virus
XX replication in cell culture or insect larvae. These genes can be deleted
XX from the genome to: (a) provide additional sites for inserting single or
XX multiple copies of foreign genes; and (b) to reduce the size of the virus
XX complementary strand relative to the polyhedrin gene. The present
XX sequence is designated ORF 152, and is on the complementary strand
XX relative to the polyhedrin gene. (Updated on 16-OCT-2003 to standardise
```

```

CC OS field)
SQ Sequence 279 BP; 88 A; 36 C; 68 G; 87 T; 0 U; 0 Other;

Query Match      18.0%; Score 101.4; DB 2; Length 279;
Best Local Similarity 75.4%; Pred. No. 2.5e-22;
Matches 126; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 112 AACATTTTGGCACTGCAAAAACACGCTTTTGGACGGGGCCCATACATAGTACAAACT 171
   |||||||
Db 276 AATTTTGGCAATGCAAAAAGTTCACTTTTGGCTGGACACCCATATACAGTACAACT 217
   |||||||

QY 172 CTACGTTTCGTAGACTATTTTACATAAATAGTCTACACGGTGTATACGCTCCAAATACA 231
   |||||||
Db 216 CTACAAATCGTAGACTATTTTATTAGAATAGTCTACACTGTACGATACGCTCCCAATATA 157
   |||||||

QY 232 CTACCACACATTTGAACCTTTTTCAGTGCACAAAAGTACGTCGCGC 278
   |||||||
Db 156 CTACTACACTATCAACTTTTTCGATTACAAAAGTTCATTTTTCG 110
   |||||||

RESULT 6
AAT13635/c
ID AAT13635 standard; DNA; 133894 BP.
XX
AC AAT13635;
DT 16-OCT-2003 (revised)
DT 03-SEP-1996 (first entry)
XX
XX AcNPV genomic DNA clone 6.
XX
XX Autographa californica nuclear polyhedrosis virus clone 6; disruption;
KW non-essential gene; heterologous protein production; expression vector;
KW baculovirus; ss.
XX
XX Autographa californica nucleopolyhedrovirus; clone 6.
XX
XX WO9601320-A2.
XX
XX 18-JAN-1996.
XX
XX 30-JUN-1995; 95WO-IB000578.
XX
XX 04-JUL-1994; 94GB-00013420.
XX
XX (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
XX
XX Bishop D, Possee R, Ayres M;
XX
XX WPI; 1996-087670/09.
XX
XX Autographa californica nuclear polyhedrosis virus complete genome
PT sequence - useful in the prodn. of vectors for enhanced heterologous
PT protein expression, such as interleukin(s), interferon(s) and
PT neurotoxin(s).
XX
XX Disclosure; Page 90-186; 122pp; English.
XX
XX The complete nucleotide sequence of the genome of clone 6 of the
CC baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV) has
CC been determined. The sequence is taken from the Genbank record L22858.
CC The patent specification claims a polynucleotide selected from open
CC reading frames (ORFs 13, 20, 22-26, 28-30, 32, 38, 41-46, 50-60, 62-63,
CC 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130, 140-146, 148-
CC 150, 152 and 154 from a total of 154 ORFs identified by the patentees.
CC See T13636-731. Expression vectors contg. the complete genomic sequence
CC of AcNPV, with the exception that at least one non-essential ORF is
CC disrupted or replaced are useful for the synthesis of heterologous
CC proteins. (Updated on 16-OCT-2003 to standardise OS field)
XX
XX Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T; 0 U; 0 Other;

Query Match      16.6%; Score 93.4; DB 10; Length 28413;
Best Local Similarity 72.5%; Pred. No. 8.6e-19;
Matches 121; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 112 AACATTTTGGCACTGCAAAAACACGCTTTTGGACGGGGCCCATACATAGTACAAACT 171
   |||||||
Db 22231 AATTTTGGCAATGCAAAAAGTTCGCTATGTTTGACATATAATATACAGTACGAACT 22172
   |||||||

QY 172 CTACGTTTCGTAGACTATTTTACATAAATAGTCTACACGGTGTATACGCTCCAAATACA 231
   |||||||
Db 22171 CTACAAATCGTAGACTATTTTATTAGAATAGTCTACACTGTACTACTACGCTCTCAATATA 22112
   |||||||

QY 232 CTACCACACATTTGAACCTTTTTCAGTGCACAAAAGTACGTCGCGC 278
   |||||||
Db 22111 CTACTACACTATCAACTTTTTCGATTACAAAAGTTCATTTTTCG 22065
   |||||||

Query Match      18.0%; Score 101.4; DB 2; Length 133894;
Best Local Similarity 75.4%; Pred. No. 4.4e-21;
Matches 126; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 112 AACATTTTGGCACTGCAAAAACACGCTTTTGGACGGGGCCCATACATAGTACAAACT 171
   |||||||
Db 132384 AATTTTGGCAATGCAAAAAGTTCACCTTTTGGCTGGACACCCATATACAGTACAACT 132325
   |||||||

QY 172 CTACGTTTCGTAGACTATTTTACATAAATAGTCTACACGGTGTATACGCTCCAAATACA 231
   |||||||
Db 132324 CTACAAATCGTAGACTATTTTATTAGAATAGTCTACACTGTACGATACGCTCCCAATATA 132265
   |||||||

QY 232 CTACCACACATTTGAACCTTTTTCAGTGCACAAAAGTACGTCGCGC 278
   |||||||
Db 132264 CTACTACACTATCAACTTTTTCGATTACAAAAGTTCATTTTTCG 132218
   |||||||

RESULT 7
ADC51646/c
ID ADC51646 standard; DNA; 28413 BP.
XX
XX ADC51646;
XX
XX 18-DEC-2003 (first entry)
XX
XX BmNPV genomic DNA nucleotides 100001-128413.
XX
XX ds; silkworm; silkworm movement suppression.
XX
XX Bombyx mori nuclear polyhedrosis virus.
XX
XX JP2003024062-A.
XX
XX 28-JAN-2003.
XX
XX 10-JUL-2001; 2001JP-00209305.
XX
XX 10-JUL-2001; 2001JP-00209305.
XX
XX (RIKA ) RIKAGAKU KENKYUSHO.
XX
XX WPI; 2003-516415/49.
XX
XX Novel Bombyx mori polyhedrosis virus in which open reading frame gene is
PT inactivated, or its variant, useful for suppressing movement of Bombyx
PT mori.
XX
XX Claim 5; SEQ ID NO 3; 53pp; Japanese.
XX
XX The invention relates to a Bombyx mori (silk worm) polyhedrosis virus in
CC which open reading frame 8 (ORF 8) gene is inactivated. The inactivated
CC orf8 gene is useful for suppressing movement of silkworm and efficiently
CC prevents the movement of a silkworm from a chamber in which it is raised.
CC The present sequence is used in the exemplification of the invention.
XX
XX Sequence 28413 BP; 8394 A; 5713 C; 5940 G; 8366 T; 0 U; 0 Other;

Query Match      16.6%; Score 93.4; DB 10; Length 28413;
Best Local Similarity 72.5%; Pred. No. 8.6e-19;
Matches 121; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 112 AACATTTTGGCACTGCAAAAACACGCTTTTGGACGGGGCCCATACATAGTACAAACT 171
   |||||||
Db 22231 AATTTTGGCAATGCAAAAAGTTCGCTATGTTTGACATATAATATACAGTACGAACT 22172
   |||||||

QY 172 CTACGTTTCGTAGACTATTTTACATAAATAGTCTACACGGTGTATACGCTCCAAATACA 231
   |||||||
Db 22171 CTACAAATCGTAGACTATTTTATTAGAATAGTCTACACTGTACTACTACGCTCTCAATATA 22112
   |||||||

QY 232 CTACCACACATTTGAACCTTTTTCAGTGCACAAAAGTACGTCGCGC 278
   |||||||
Db 22111 CTACTACACTATCAACTTTTTCGATTACAAAAGTTCATTTTTCG 22065
   |||||||

```

```
RESULT 8
ADQ48576
ID ADQ48576 standard; DNA; 141 BP.
XX
XX ADQ48576;
XX
XX 09-SEP-2004 (first entry)
XX
XX Viral vector-related plasmid pIB/V5-His-DEST recombination region #1.
XX
XX viral vector; recombination site; recombinant virus;
XX replication-defective particle generation; gene expression inhibition;
XX gene therapy vector; ds; plasmid; recombination region.
XX
XX Unidentified.
XX
XX WO2004009768-A2.
XX
XX 29-JAN-2004.
XX
XX 18-JUL-2003; 2003WO-US022437.
XX
XX 18-JUL-2002; 2002US-0396335P.
XX PR 26-JUL-2002; 2002US-0398617P.
XX PR 19-NOV-2002; 2002US-0427231P.
XX PR 24-MAR-2003; 2003US-0456496P.
XX PR 03-JUN-2003; 2003US-0474940P.
XX
XX (INVI-) INVITROGEN CORP.
XX PA (BENN/) BENNETT R P.
XX PA (WELC/) WELCH P J.
XX PA (HARW/) HARWOOD S.
XX PA (MADD/) MADDEN K.
XX PA (FRIM/) FRIMPONG K.
XX PA (FRAN/) FRANKE K E.
XX
XX Bennett RP, Welch PJ, Harwood S, Madden K, Frimpong K, Franke KE;
XX WPI; 2004-132944/13.
XX
XX Novel nucleic acid molecule comprising all or a portion of one or more
XX viral genome and further comprising two or more recombination sites that
XX do not substantially recombine with each other, useful as gene therapy.
XX
XX Disclosure; Fig 17; 555pp; English.
XX
XX The invention comprises a nucleic acid molecule consisting of all or a
XX portion of one or more viral/baculoviral genomes and further containing
XX at least two recombination sites that do not substantially recombine with
XX each other. One or more of the recombination sites is capable of
XX undergoing recombination with a compatible recombination site in the
XX presence of one or more proteins active in lambda recombination. The
XX nucleic acid of the invention replicates in prokaryotic and eukaryotic
XX cells. The nucleic acid of the invention is useful for constructing a
XX recombinant virus, generating replication-defective particles,
XX preventing/inhibiting the expression of one or more genes in an organism,
XX and are useful as gene therapy vectors. The nucleic acid of the invention
XX is also useful for producing and expressing fusion polypeptides. The
XX present DNA sequence represents a recombination region of a plasmid that
XX was used in the exemplification of the invention.
XX
XX Sequence 141 BP; 43 A; 36 C; 27 G; 34 T; 0 U; 1 Other;
XX
XX Query Match 11.0%; Score 62; DB 12; Length 141;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-09;
XX Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 487 CTTATCGCGCCTATAATACAGCCGCAACGATCTGTGGTAAACAGAGTTGAACAGCATCTG 546
XX
XX 1 CTTATCGCGCCTATAATACAGCCGCAACGATCTGTGGTAAACAGAGTTGAACAGCATCTG 60
XX
XX 547 TT 548
XX
XX Db
XX 61 TT 62
XX
XX RESULT 9
XX ABZ10059/c
XX ID ABZ10059 standard; DNA; 6289 BP.
XX
XX ABZ10059;
XX
XX 16-JAN-2003 (first entry)
XX
XX Haematopoietic cell proliferation disorder related DNA sequence #199.
XX
XX Human; haematopoietic cell proliferation disorder; cytostatic;
XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
XX cytosine methylation state; gene; ds.
XX
XX Homo sapiens.
XX
XX WO20027272-A2.
XX
XX 03-OCT-2002.
XX
XX 26-MAR-2002; 2002WO-EP003401.
XX
XX 26-MAR-2001; 2001US-0278333P.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
XX Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
XX Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
XX Schwobe I, Ziebarth H;
XX WPI; 2003-018942/01.
XX
XX Detecting and differentiating between hematopoietic cell proliferative
XX disorders, comprises contacting a target nucleic acid with a reagent that
XX distinguishes between methylated and non-methylated CpG dinucleotides.
XX
XX Claim 28; SEQ ID NO 199; 117pp; English.
XX
XX The present invention describes a method for detecting and
XX differentiating between haematopoietic cell proliferative disorders
XX associated with at least 1 gene and/or their regulatory regions in a
XX subject. The method comprises contacting a target nucleic acid in a
XX biological sample obtained from the subject with at least 1 reagent,
XX which distinguishes between methylated and non-methylated CpG
XX dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
XX represent specifically claimed nucleotide sequences from the present
XX invention. Oligonucleotides from the present invention can be used: for
XX differentiating between healthy haematopoietic cells and proliferative
XX disorder haematopoietic cells; for differentiating between acute
XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
XX determining the cytosine methylation state and/or single nucleotide
XX polymorphisms (SNPs) of haematopoietic cell proliferation disorder
XX related sequences and their complements; and as primers for the
XX amplification of haematopoietic cell proliferation disorder related DNA
XX sequences. The nucleotide sequences from the present invention can also
XX be used for detecting a predisposition to, differentiation between
XX subclasses, diagnosis, prognosis, treatment and/or monitoring of
XX haematopoietic cell proliferative disorders. The present method enables a
XX highly specific classification of haematopoietic cell proliferative
XX disorders allowing for improved and informed treatment of patients
XX
XX Sequence 6289 BP; 1680 A; 178 C; 1488 G; 2943 T; 0 U; 0 Other;
XX
XX Query Match 7.0%; Score 39.6; DB 8; Length 6289;
XX Best Local Similarity 53.2%; Pred. No. 0.14;
XX Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
XX
XX 112 AACTTTTTCGACTGCAGAAAAAACACGCTTTTTCACGCGGGCCCATACATAGTACAAACT 171
XX
```

Db 4896 AATTTATTTATTAACCAATAAACTATTATATAATCCACTACCAATAATCATTCACAAA 4837
QY 172 CTACGTTTCGTAGACTATTTTACATAAAATAGTCTACACGGTTGTATACGCTCCAAATACA 231
Db 4836 TAATTAATTTTAAACATTTATTATATAAAATCGTCACAAATCTTTTACTCAAAAATA 4777
QY 232 CTACACACATTTGACCTTTTTCAGTGCACAAAAGTA 269
Db 4776 ATACAACCTTCTCTCCCTATATAAAACGAAAAAATA 4739

RESULT 10
AAS46501/c
ID AAS46501 standard; DNA; 9289 BP.
XX AC AAS46501;
XX DT 18-DEC-2001 (first entry)
XX DE Tumour suppressor gene derived chemically modified sequence #223.
XX KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX OS Homo sapiens.
XX PN WO200168912-A2.
XX PD 20-SEP-2001.
XX PF 15-MAR-2001; 2001WO-BP002955.
XX PR 15-MAR-2000; 2000DE-01013847.
XX PR 06-APR-2000; 2000DE-01019058.
XX PR 07-APR-2000; 2000DE-01019173.
XX PR 30-JUN-2000; 2000DE-01032529.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602752/68.
XX DR Fragments of chemically modified genes associated with tumor suppressor
XX genes and oncogenes, useful in designing primers and probes for analyzing
XX diseases associated with cytosine methylation state e.g. cancer.
XX PS Claim 1; SEQ ID NO 223; 27pp; English.
XX CC The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and oncogenes
XX having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX 500 are missing from the sequence listing) sequences (SS) and sequences
XX complementary to (SS). The nucleic acid may be a peptide nucleic acid-
XX oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX probes for detecting the cytosine methylation state and/or single
XX nucleotide polymorphisms and also to be used in an array for analysing
XX diseases associated with CpG dinucleotides e.g. cancers and tumours. The
XX probes can also be used in a method for ascertaining genetic and/or
XX epigenetic parameters for the diagnosis and/or therapy of existing
XX diseases or the predisposition to specific diseases, by analysing
XX cytosine methylations. The parameters may be compared to another set of
XX genetic and/or epigenetic parameters, the differences serving as basis
XX for diagnosis and/or prognosis events which are disadvantageous to
XX patients. The present sequence is one of the 533 genomic sequences
XX derived from tumour suppressor genes and oncogenes. Note: The sequence
XX data for this patent did not form part of the printed specification, but
XX was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 9289 BP; 2515 A; 206 C; 2097 G; 4471 T; 0 U; 0 Other;
Query Match 7.0%; Score 39.6; DB 4; Length 9289;
Best Local Similarity 53.2%; Pred. No. 0.16; Mismatches 0; Gaps 0;
Matches 84; Conservative 0; Indels 74; Indels 0; Gaps 0;
QY 112 AACTTTTTTGCACTGCAAAAAACACGCTTTTTCAGCGGGCCCATACATAGTACAAACT 171
Db 7896 AATTTATTTATATAACCAATAAACTATTATATAATCCACTACCAATAATCATTCACAAA 7837
QY 172 CTACGTTTCGTAGACTATTTTACATAAAATAGTCTACACGGTTGTATACGCTCCAAATACA 231
Db 7836 TAATTAATTTTAAACATTTATTATATAAAATCGTCACAAATCTTTTACTCAAAAATA 7777
QY 232 CTACACACATTTGACCTTTTTCAGTGCACAAAAGTA 269
Db 7776 ATACAACCTTCTCTCCCTATATAAAACGAAAAAATA 7739

RESULT 11
ADE84121/c
ID ADE84121 standard; DNA; 9289 BP.
XX AC ADE84121;
XX DT 29-JAN-2004 (first entry)
XX DE Human lymphoid cell proliferative disorder gene derived DNA #57.
XX KW ds; lymphoid cell proliferative disorder; methylation;
KW methylated CpG dinucleotide; single nucleotide polymorphism; SNP;
KW diffuse large B-cell lymphoma; mantle cell lymphoma;
KW chronic lymphocytic leukemia; small lymphocytic lymphoma;
KW follicular lymphoma; diagnosis; prognosis.
XX OS Homo sapiens.
XX PN WO2003044226-A2.
XX PD 30-MAY-2003.
XX PF 25-NOV-2002; 2002WO-EP013285.
XX PR 23-NOV-2001; 2001DE-01057491.
XX PR 28-DEC-2001; 2001DE-01064501.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Burger M, Caldwell C, Genc B, Becker E, Maier S, Nimmrich I;
XX WPI; 2003-457621/43.
XX DR Detecting and differentiating between lymphoid cell proliferative
XX disorders comprises contacting a target nucleic acid with at least one
XX reagent that distinguishes between methylated and non-methylated CpG
XX dinucleotides.
XX PS Claim 26; SEQ ID NO 117; 448pp; English.
XX CC The invention relates to a method of detecting and differentiating
XX between lymphoid cell proliferative disorders associated with at least
XX one gene and/or their regulatory regions in a subject by contacting a
XX target nucleic acid in a biological sample obtained from the subject with
XX at least one reagent or series of reagents that distinguish between
XX methylated and non-methylated CpG dinucleotides within the target nucleic
XX acid. The genes and/or their regulatory regions are preferably selected
XX from MDRI, CSNK2B, EGR4, AR, CDK4, RE2, CDC35A, GPII beta, MYOD1, CDH3,
XX MYCL1, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2a, FOS,
XX GSTP1, HIC-1, MGMT, MLH1, MOS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKNIC,
XX GSK3beta, ESRI, APAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic
XX acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
XX of the genes are useful for detecting the methylation state of all the

PT reagent that distinguishes between methylated and non-methylated CpG
 XX dinucleotides.
 PS Claim 26; SEQ ID NO 193; 448pp; English.
 XX
 CC The invention relates to a method of detecting and differentiating
 CC between lymphoid cell proliferative disorders associated with at least
 CC one gene and/or their regulatory regions in a subject by contacting a
 CC target nucleic acid in a biological sample obtained from the subject with
 CC at least one reagent or series of reagents that distinguish between
 CC methylated and non-methylated CpG dinucleotides within the target nucleic
 CC acid. The genes and/or their regulatory regions are preferably selected
 CC from MBR1, CSNK2B, EGR4, AR, CDK4, RB2, CDC45A, GPR12, MYO10, CDH3,
 CC MYC1, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2A, CDKN2B, FOS,
 CC GSTP1, HIC-1, MGMT, MLH1, MOS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKN1C,
 CC GSK3beta, ESRI, APAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic
 CC acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
 CC of the genes are useful for detecting the methylation state of all the
 CC CpG dinucleotides within one or more the sequences, or their complements,
 CC for determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs), and for differentiating at least two of the medical
 CC conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,
 CC chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular
 CC lymphoma. They are also useful for detecting of a predisposition to,
 CC differentiation between subclasses, diagnosis, prognosis, treating and/or
 CC monitoring of lymphoid cell proliferative disorder. This sequence
 CC represents a nucleic acid of a pretreated genomic DNA derived from the
 CC above mentioned genes.
 XX
 SQ Sequence 9289 BP; 2515 A; 0 C; 2097 G; 4677 T; 0 U; 0 Other;
 Query Match 6.7%; Score 38; DB 10; Length 9289;
 Best Local Similarity 52.5%; Pred. No. 0.54;
 Matches 83; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 112 AACTTTTTCGACTGCAAAAAACACGCTTTTGACGGGGCCCATACATAGTACAACT 171
 DB 7896 AATTATTATTATATAACCAATAAACTATTATAAATCCACTACCAATATCAATCAAAAA 7837
 QY 172 CTAAGTTTCGTAGACTATTATTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACA 231
 DB 7836 TAATAATTTTAAACATATTATTATAAAAAATCATCAAACTCTTTTACTCAAAAAATA 7777
 QY 232 CTAACACACATTGAACCTTTTTCGAGTGCAAAAAAGTA 269
 DB 7776 ATACAACTTTCCTCTCCCTATATAAAAAACAAAAATA 7739
 RESULT 14
 ADA71938/c
 ID ADA71938 standard; DNA; 2000 BP.
 XX
 AC ADA71938;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Rice gene, SEQ ID 5263.
 XX
 KW Plant; bacterial infection; fungal infection; viral infection; rice;
 KW gene; ds.
 XX
 OS Oryza sativa.
 XX
 PN WO2003000898-A1.
 XX
 PD 03-JAN-2003.
 XX
 PF 22-JUN-2001; 2001WO-IB001105.
 XX
 PR 22-JUN-2001; 2001WO-IB001105.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX WPI; 2003-175290/17.
 XX
 PT Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.
 XX
 PS Claim 27; SEQ ID NO 5263; 899pp; English.
 XX
 CC The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.
 XX
 SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
 Query Match 6.4%; Score 36.2; DB 8; Length 2000;
 Best Local Similarity 11.3%; Pred. No. 1;
 Matches 42; Conservative 169; Mismatches 157; Indels 3; Gaps 1;
 QY 134 CTGCAAAAAACACGCTTTTGACGGGGCCCATACATAGTACAACTACGTTTCGTA 183
 DB 430 CYGCKWMTYCYGYMKWYTMGSYKYSRYCYKMYRMWYKGMWYMYYSAYSSMTWYYY 371
 QY 184 GACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACACACATT 243
 DB 370 AKYWKYWKYRGTMSWYKSYKKYC---TWYCNKMKRCYRWKMKMKTKYSKRCYCW 314
 QY 244 GAACCTTTTTCAGTGCAAAAAAGTACGTGTGGCAGTACGTCAGCGCGGCGCTTATCGGG 303
 DB 313 RVATCYWCCYKRGWYSRRSMRTAGWKWRSWRSWRSYKWKYKMKWKKSYMSYG 254
 QY 304 TCGCGTCTCTGCACGTACGAATCACTATTCCGACCGGACGAGCTGTTCCTTATCGTGAC 363
 DB 253 WARSSGTWSRSAAKRTYKGYSTSRRAKMKRACMYSAACRRYSRTSYCCSCYCGSSKWKY 194
 QY 364 AGGACGCCAGCTTCTGTGTCTAAACCGCAGCGGACGCAACTCTTATCGCAACAGGA 423
 DB 193 MSKSCSMRTCSSWCSCCYTCYGGAMCWSCCMSWYMGSCGCTTRGWKRSKYSCKKY 134
 QY 424 CGCGCTCCATATACAGCGCGGTTATCTCATGCGGTGACCGGACACGAGCGCGCGTC 483
 DB 133 CSCCTKYCSYTGYYRYCKWKYKYSYKCYCYWYMSYMYMKWCMKCSRSCSSWMSAYC 74
 QY 484 CCGCTTATCGC 494
 DB 73 STSSTRWMSM 63
 RESULT 15
 ABQ68715/c
 ID ABQ68715 standard; DNA; 986 BP.
 XX
 AC ABQ68715;
 XX
 DT 29-AUG-2003 (revised)
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria monocytogenes 4b contig DNA sequence #1481.
 XX
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 KW infection; ds.
 XX
 OS Listeria monocytogenes ATCC 19115.

```
XX WO200228891-A2.
XX
XX
PD 11-APR-2002.
XX
XX 04-OCT-2001; 2001WO-FR003061.
XX
XX 04-OCT-2000; 2000FR-00012697.
XX
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Kunst F, Glaser P;
XX
XX WPI; 2002-332479/37.
XX
XX New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX antibodies and modulators.
XX
XX Claim 14; SEQ ID NO 1528; 180pp; French.
XX
XX The present invention relates to nucleic acid sequences (ABQ67188-
XX ABQ71212) from Listeria sp. The sequences are useful as probes and
XX primers for identification and/or detection of Listeria (e.g. as
XX contaminants in foods, or mutational analysis) and for analysis of gene
XX expression. Proteins encoded by the nucleic acid sequences can be used to
XX screen for compounds that modulate gene expression, replication and
XX pathogenicity of Listeria (potential therapeutic agents), also for
XX treating infections by Listeria, and are useful as immunogens in anti-
XX Listeria vaccines. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
XX on 29-AUG-2003 to standardise OS field)
XX
XX SQ Sequence 986 BP; 295 A; 199 C; 157 G; 328 T; 0 U; 7 Other;
XX
XX Query Match 6.3%; Score 35.4; DB 6; Length 986;
XX Best Local Similarity 56.4%; Pred. No. 1.3;
XX Matches 66; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
XX
XX Qy 155 CATACATAGTACAAACTCTACGTTTCGTGAGACTATTTTACATAAATAGTCTACACCGTTG 214
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 318 CAGATTTAGAAAACATTTCGTTGCCCGAGAAAATTTAAATATCAAGAGCACTCTGTTG 259
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 215 TATACGCTCCAAATACACTACACACATTGAAACCTTTTTCGAGTGCAGAAAAAGTACG 271
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 258 CAGAACATTCATATAAAGTTACATCCATTGCTCAATTTTCGGTGCTGTAGAGGAAG 202
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: October 24, 2005, 18:57:59
Job time : 624.051 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 20:18:24 ; Search time 1150.82 Seconds
(without alignments)
4044.488 Million cell updates/sec

Title: US-09-896-888A-1

Perfect score: 564

Sequence: 1 catgatgataacaatgtat.....tggtagcgacacacatg 564

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9772363 seqs, 4126298632 residues

Total number of hits satisfying chosen parameters: 19544726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/ECT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/ECTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	564	100.0	564	US-09-896-888A-1	Sequence 1, Appli
2	548	97.2	560	US-10-622-088-126	Sequence 126, App
3	548	97.2	2773	US-10-295-074-60	Sequence 60, Appl
4	548	97.2	2773	US-10-846-911-60	Sequence 60, Appl
5	548	97.2	2773	US-10-939-107-60	Sequence 60, Appl

6	546.4	96.9	5038	21	US-10-622-088-89	Sequence 89, Appl
7	462	81.9	462	9	US-09-896-888A-14	Sequence 14, Appl
8	62	11.0	47	21	US-10-622-088-127	Sequence 127, App
9	62	11.0	325	21	US-10-622-088-149	Sequence 149, App
10	57	10.1	92	9	US-09-896-888A-16	Sequence 16, Appl
11	42.6	7.6	88	9	US-09-896-888A-15	Sequence 15, Appl
12	39.6	7.0	6289	21	US-10-473-126-199	Sequence 199, App
13	39.6	7.0	9289	19	US-10-221-714A-223	Sequence 223, App
14	38	6.7	6289	21	US-10-473-126-345	Sequence 345, App
15	35.4	6.3	986	18	US-10-398-221-1528	Sequence 1528, App
16	35.4	6.3	1549	18	US-10-398-221-3152	Sequence 3152, App
17	34.4	6.1	855	9	US-09-981-876-89	Sequence 89, Appl
18	34.4	6.1	855	10	US-09-148-545-89	Sequence 89, Appl
19	34.4	6.1	855	24	US-10-979-111-89	Sequence 89, Appl
20	34.2	6.1	436	21	US-10-425-115-104158	Sequence 104158,
21	34	6.0	18977	26	US-11-097-143-26311	Sequence 26311, A
22	33.8	6.0	600	26	US-11-060-756-2178	Sequence 2178, Ap
23	33.8	6.0	600	26	US-11-060-756-2179	Sequence 2179, Ap
24	33.8	6.0	600	26	US-11-060-756-6450	Sequence 6450, Ap
25	33.8	6.0	600	26	US-11-060-756-6451	Sequence 6451, Ap
26	33.8	6.0	930	13	US-09-925-065A-27884	Sequence 27884, A
27	33.8	6.0	1726	13	US-09-925-065A-68947	Sequence 68947, A
28	33.8	6.0	2117	14	US-10-087-192-1463	Sequence 1463, Ap
29	33.8	6.0	3056	16	US-10-225-567A-125	Sequence 125, App
30	33.8	6.0	3083	17	US-10-166-568-1	Sequence 1, Appli
31	33.8	6.0	39344	14	US-10-087-192-1462	Sequence 1462, Ap
32	33.2	5.9	573	13	US-09-925-065A-559402	Sequence 559402,
33	33.2	5.9	621	20	US-10-437-963-77138	Sequence 77138, A
34	33	5.9	14429	16	US-10-311-455-2215	Sequence 2215, Ap
35	33	5.9	14429	20	US-10-433-793-127	Sequence 127, App
36	32.8	5.8	426	24	US-10-779-543-9174	Sequence 9174, Ap
37	32.8	5.8	454	9	US-09-770-444-580	Sequence 580, App
38	32.8	5.8	2000	9	US-09-938-842A-5148	Sequence 5148, Ap
39	32.8	5.8	2000	11	US-09-938-842A-5148	Sequence 5148, Ap
40	32.6	5.8	469	21	US-10-425-115-102790	Sequence 102790,
41	32.6	5.8	495	10	US-09-814-353-5212	Sequence 5212, Ap
42	32.6	5.8	495	10	US-09-814-353-11499	Sequence 11499, A
43	32.6	5.8	535	13	US-09-925-065A-493593	Sequence 493593,
44	32.6	5.8	535	13	US-09-925-065A-493600	Sequence 493600,
45	32.6	5.8	600	13	US-09-925-065A-493597	Sequence 493597,

ALIGNMENTS

RESULT 1
US-09-896-888A-1
; Sequence 1, Application US/09896888A
; Patent No. US20020116723A1
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Insect Expression Vectors
; FILE REFERENCE: 80021-44
; CURRENT APPLICATION NUMBER: US/09/896,888A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US/09/048,911
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/049,946
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Orgyia pseudotsugata
US-09-896-888A-1

Query Match 100.0%; Score 564; DB 9; Length 564;
Best Local Similarity 100.0%; Pred. No. 3.3e-175;
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGATGATAACAATGTATGCTGCTTAATGCTTCAACAACAAATTCGTTGAACGTG 60
|||||

Db 1 CATGATGATAAACAATGATGTCGCTAATGTTGCTTCAACAACAATCTGTTGAACGTG 60
Qy 61 TTTTCATGTTTGGCCCAAGACCTTTTATCTCGGTGGCCTCCCGACCAACTTTTTT 120
Db 61 TTTTCATGTTTGGCCCAAGACCTTTTATCTCGGTGGCCTCCCGACCAACTTTTTT 120
Qy 121 GCACGTGCAAAAAACACCGCTTTTGACCGGGCCCATACATAGTACAACTCTAGTTTC 180
Db 121 GCACGTGCAAAAAACACCGCTTTTGACCGGGCCCATACATAGTACAACTCTAGTTTC 180
Qy 181 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTTCCAAATACACTACCAAC 240
Db 181 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTTCCAAATACACTACCAAC 240
Qy 241 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGCGGAGTCAGTAGGCGGCTTATC 300
Db 241 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGCGGAGTCAGTAGGCGGCTTATC 300
Qy 301 GGGTCGCGCTCTGTCAACCTACGAATCACAATTATCGGACCGGACGAGTGTCTTATCGT 360
Db 301 GGGTCGCGCTCTGTCAACCTACGAATCACAATTATCGGACCGGACGAGTGTCTTATCGT 360
Qy 361 GACAGAGCGCAGCTTCTGTGTTGCTAAACCGAGCGGACGCAACTCTTATCGGAACA 420
Db 361 GACAGAGCGCAGCTTCTGTGTTGCTAAACCGAGCGGACGCAACTCTTATCGGAACA 420
Qy 421 GGACGGCGCTCCATATACGCGCGGCTTATCTCATGCGCGTGACCGGACAGGCGGCC 480
Db 421 GGACGGCGCTCCATATACGCGCGGCTTATCTCATGCGCGTGACCGGACAGGCGGCC 480
Qy 481 GTCCCGCTTATCGCGCTTATAAATACGCGCGCAACGATCTGGTAAACACAGTTGAACAG 540
Db 481 GTCCCGCTTATCGCGCTTATAAATACGCGCGCAACGATCTGGTAAACACAGTTGAACAG 540
Qy 541 CATCTGTTACAGCGACACAACATG 564
Db 541 CATCTGTTACAGCGACACAACATG 564

RESULT 2

US-10-622-088-126
; Sequence 126, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth
; APPLICANT: Franke, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622,088
; PRIORITY FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/22437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 126
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:

; OTHER INFORMATION: OpIE2 promoter sequence
US-10-622-088-126

Query Match 97.2%; Score 548; DB 21; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.3e-170;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGATGATAAACAATGATGTCGCTAATGTTGCTTCAACAACAATCTGTTGAACGTG 60
Db 5 CATGATGATAAACAATGATGTCGCTAATGTTGCTTCAACAACAATCTGTTGAACGTG 64
Qy 61 TTTTCATGTTTGGCCCAAGACCTTTTATCTCGGTGGCCTCCCGACCAACTTTTTT 120
Db 65 TTTTCATGTTTGGCCCAAGACCTTTTATCTCGGTGGCCTCCCGACCAACTTTTTT 124
Qy 121 GCACGTGCAAAAAACACCGCTTTTGACCGGGCCCATACATAGTACAACTCTAGTTTC 180
Db 125 GCACGTGCAAAAAACACCGCTTTTGACCGGGCCCATACATAGTACAACTCTAGTTTC 184
Qy 181 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTTCCAAATACACTACCAAC 240
Db 185 GTAGACTATTTTACATAAATAGTCTACACCGTTGTATAGCTTCCAAATACACTACCAAC 244
Qy 241 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGCGGAGTCAGTAGGCGGCTTATC 300
Db 245 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGCGGAGTCAGTAGGCGGCTTATC 304
Qy 301 GGGTCGCGCTCTGTCAACCTACGAATCACAATTATCGGACCGGACGAGTGTCTTATCGT 360
Db 305 GGGTCGCGCTCTGTCAACCTACGAATCACAATTATCGGACCGGACGAGTGTCTTATCGT 364
Qy 361 GACAGGAGCGCAGCTTCTGTGTTGCTAAACCGAGCGGACGCAACTCTTATCGGAACA 420
Db 365 GACAGGAGCGCAGCTTCTGTGTTGCTAAACCGAGCGGACGCAACTCTTATCGGAACA 424
Qy 421 GGACGGCGCTCCATATACGCGCGGCTTATCTCATGCGCGTGACCGGACAGGCGGCC 480
Db 425 GGACGGCGCTCCATATACGCGCGGCTTATCTCATGCGCGTGACCGGACAGGCGGCC 484
Qy 481 GTCCCGCTTATCGCGCTTATAAATACGCGCGCAACGATCTGGTAAACACAGTTGAACAG 540
Db 485 GTCCCGCTTATCGCGCTTATAAATACGCGCGCAACGATCTGGTAAACACAGTTGAACAG 544
Qy 541 CATCTGTT 548
Db 545 CATCTGTT 552

RESULT 3

US-10-295-074-60
; Sequence 60, Application US/10295074
; Publication No. US20030185845A1
; GENERAL INFORMATION:
; APPLICANT: Pharmexa A/S
; TITLE OF INVENTION: NOVEL IMMUNOGENIC MIMETICS OF MULTIMER PROTEINS
; FILE REFERENCE: P1013DK00
; CURRENT APPLICATION NUMBER: US/10/295,074
; CURRENT FILING DATE: 2002-11-15
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 2773
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: p22Op2F expression vector for insect cells
; FEATURE:
; NAME/KEY: misc recomb
; LOCATION: (561)..(566)
; OTHER INFORMATION: HindIII site
; FEATURE:
; NAME/KEY: misc recomb
; LOCATION: (573)..(578)

OTHER INFORMATION: AvalI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (586)..(591)
OTHER INFORMATION: EcoRI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (593)..(598)
OTHER INFORMATION: BamHI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (625)..(630)
OTHER INFORMATION: ClaI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (629)..(634)
OTHER INFORMATION: ClaI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (1156)..(1161)
OTHER INFORMATION: ApaLI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2128)..(2133)
OTHER INFORMATION: PstI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2204)..(2209)
OTHER INFORMATION: NcoI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2284)..(2289)
OTHER INFORMATION: AvalI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2294)..(2299)
OTHER INFORMATION: AvalI, SmaI, and XmaI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2551)..(2556)
OTHER INFORMATION: ApaLI site

US-10-295-074-60

Query Match 97.2%; Score 548; DB 17; Length 2773;

Best Local Similarity 100.0%; Pred. No. 1.4e-169;

Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGATGATAACAATGATGTGCTAATGTTGCTTCAACAACAATCTGTGAACTGTG 60
DB |||||||
QY 5 CATGATGATAACAATGATGTGCTAATGTTGCTTCAACAACAATCTGTGAACTGTG 64
DB |||||||
QY 61 TTTTCATGTTGCCCAACAGCACCTTTTACTCGGTGGCTCCGCCACCACTTTT 120
DB |||||||
QY 65 TTTTCATGTTGCCCAACAGCACCTTTTACTCGGTGGCTCCGCCACCACTTTT 124
DB |||||||
QY 121 GCACGTGCAAAAACACGCTTTTGACCGGGCCCATACATAGTACAACTCTAGTTTC 180
DB |||||||
QY 125 GCACGTGCAAAAACACGCTTTTGACCGGGCCCATACATAGTACAACTCTAGTTTC 184
DB |||||||
QY 181 GTAGACTATTTTACATAAATAGTCTACACCGTGTATAGCTCCAAATACACTACAC 240
DB |||||||
QY 185 GTAGACTATTTTACATAAATAGTCTACACCGTGTATAGCTCCAAATACACTACAC 244
DB |||||||
QY 241 ATTGAACCTTTTTCAGTGTCAAAAAGTACGTGTGGCAGTCACTAGCGCGCTTATC 300
DB |||||||
QY 245 ATTGAACCTTTTTCAGTGTCAAAAAGTACGTGTGGCAGTCACTAGCGCGCTTATC 304
DB |||||||
QY 301 GGGTCGCGCTCTGTCAACGTACGAATCACAATTACGGACCGGACGAGTGTGTCTTATCGT 360
DB |||||||
QY 305 GGGTCGCGCTCTGTCAACGTACGAATCACAATTACGGACCGGACGAGTGTGTCTTATCGT 364
DB |||||||
QY 361 GACAGGACGCGAGCTTCTGTGTTGCTTAACCGCAGCGGACGCAACTCTCTTATCGGAACA 420
DB |||||||

Db 365 GACAGGACGCGAGCTTCTGTGTTGTTAAACGCGAGCGGACGCAACTCTCTTATCGGAACA 424
QY 421 GGACGCGCTCCATATCAGCCGCGGTTTATCTCATGCGCGTGACCGGACACGAGCGGCC 480
DB |||||||
Db 425 GGACGCGCTCCATATCAGCCGCGGTTTATCTCATGCGCGTGACCGGACACGAGCGGCC 484
QY 481 GTCCGCGTTATCGCGCTTATAAATACAGCCGCGCAACGATCTGGTAAACACAGTTGAACAG 540
DB |||||||
Db 485 GTCCGCGTTATCGCGCTTATAAATACAGCCGCGCAACGATCTGGTAAACACAGTTGAACAG 544
QY 541 CATCTGTT 548
DB 545 CATCTGTT 552

RESULT 4
US-10-846-911-60
; Sequence 60, Application US/10846911
; Publication No. US20040258660A1
; GENERAL INFORMATION:
; APPLICANT: KLYSNER, Steen
; APPLICANT: NIELSEN, Finn Stausholm
; APPLICANT: BRATT, Tomas
; APPLICANT: VOLDORGB, Bjorn
; APPLICANT: MOURITSEN, Soren
; TITLE OF INVENTION: NOVEL IMMUNOGENIC MIMETICS OF MULTIMER PROTEINS
; FILE REFERENCE: 674542-2018
; CURRENT APPLICATION NUMBER: US/10/846,911
; CURRENT FILING DATE: 2004-05-14
; PRIOR APPLICATION NUMBER: PCT/DK02/00764
; PRIOR FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: 60/331,575
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: PA 2001 01702
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 2773
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: p2Op2F expression vector for insect cells
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (561)..(566)
; OTHER INFORMATION: HindIII site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (573)..(578)
; OTHER INFORMATION: AvalI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (586)..(591)
; OTHER INFORMATION: EcoRI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (593)..(598)
; OTHER INFORMATION: BamHI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (625)..(630)
; OTHER INFORMATION: ClaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (629)..(634)
; OTHER INFORMATION: ClaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (1156)..(1161)
; OTHER INFORMATION: ApaLI site
; FEATURE:
; NAME/KEY: misc_recomb

; LOCATION: (2128)..(2133)
; OTHER INFORMATION: PstI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2204)..(2209)
; OTHER INFORMATION: NcoI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2284)..(2289)
; OTHER INFORMATION: AvaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2294)..(2299)
; OTHER INFORMATION: AvaI, SmaI, and XmaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2551)..(2556)
; OTHER INFORMATION: ApaLI site
US-10-846-911-60

Query Match 97.2%; Score 548; DB 21; Length 2773;
Best Local Similarity 100.0%; Pred. No. 1.4e-169;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CATGATGATAAACAATGATGGTGCTAATGTTGCTTCAACAACAATTCGTGTGAACGTG 60
Db |||||||
Qy 5 CATGATGATAAACAATGATGGTGCTAATGTTGCTTCAACAACAATTCGTGTGAACGTG 64
Db |||||||
Qy 61 TTTTCATGTTTGCACAAGACCTTTATATCTCGGTGGCCTCCCAACCAACTTTTTT 120
Db |||||||
Qy 65 TTTTCATGTTTGCACAAGACCTTTATATCTCGGTGGCCTCCCAACCAACTTTTTT 124
Db |||||||
Qy 121 GCACTGCAAAAAACACGCTTTTCACCGCGGCCCATACATAGTACAACTCTACGTTTC 180
Db |||||||
Qy 125 GCACTGCAAAAAACACGCTTTTCACCGCGGCCCATACATAGTACAACTCTACGTTTC 184
Db |||||||
Qy 181 GTAGACTATTTTACATAAATAGTCTACACCGTGTGTATACGCTCCAAATACACTACCAAC 240
Db |||||||
Qy 185 GTAGACTATTTTACATAAATAGTCTACACCGTGTGTATACGCTCCAAATACACTACCAAC 244
Db |||||||
Qy 241 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGCGGAGTCAAGTGAGCGCGGCTTATC 300
Db |||||||
Qy 245 ATTGAACCTTTTTCAGTGCAAAAAGTACGTGCGGAGTCAAGTGAGCGCGGCTTATC 304
Db |||||||
Qy 301 GGGTCGGCTCTGTACGTACGATTCATATCGGACCGGACGAGTGTGCTTTATCGT 360
Db |||||||
Qy 305 GGGTCGGCTCTGTACGTACGATTCATATCGGACCGGACGAGTGTGCTTTATCGT 364
Db |||||||
Qy 361 GACAGGACCGCAGCTTCTGTGTTGCTAAACCGCAGCGGACCACTCTTATCGGAACA 420
Db |||||||
Qy 365 GACAGGACCGCAGCTTCTGTGTTGCTAAACCGCAGCGGACCACTCTTATCGGAACA 424
Db |||||||
Qy 421 GGACGGCCTCCATATACGCGCGGTTATCTATGCGCGTGACCGGACACGAGCGGCC 480
Db |||||||
Qy 425 GGACGGCCTCCATATACGCGCGGTTATCTATGCGCGTGACCGGACACGAGCGGCC 484
Db |||||||
Qy 481 GTCCGCTTATCGCGCTTAAATATACAGCCGCGCAACGATCTGGTAAACAGTTGAACAG 540
Db |||||||
Qy 485 GTCCGCTTATCGCGCTTAAATATACAGCCGCGCAACGATCTGGTAAACAGTTGAACAG 544
Db |||||||
Qy 541 CATCTGTT 548
Db |||||||
Qy 545 CATCTGTT 552
Db |||||||

RESULT 5

US-10-939-107-60
; Sequence 60, Application US/10939107
; Publication No. US20050180947A1
; GENERAL INFORMATION:
; APPLICANT: Pedersen, Hans Rudolf
; APPLICANT: Ebert, Bjarke
; APPLICANT: Pedersen, Louise Henriette

; APPLICANT: Rasmussen, Peter Birk
; TITLE OF INVENTION: Novel Application of Vaccination Against TNF-alpha
; FILE REFERENCE: 674542-2020
; CURRENT APPLICATION NUMBER: US/10/939,107
; CURRENT FILING DATE: 2004-09-10
; PRIOR APPLICATION NUMBER: PCT/DK03/00147
; PRIOR FILING DATE: 2003-03-11
; PRIOR APPLICATION NUMBER: 60/363,128
; PRIOR FILING DATE: 2002-03-11
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 2773
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: p2Op2F expression vector for insect cells
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (561)..(566)
; OTHER INFORMATION: HindIII site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (573)..(578)
; OTHER INFORMATION: AvaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (586)..(591)
; OTHER INFORMATION: EcoRI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (593)..(598)
; OTHER INFORMATION: BamHI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (625)..(630)
; OTHER INFORMATION: ClaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (629)..(634)
; OTHER INFORMATION: ClaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (1156)..(1161)
; OTHER INFORMATION: ApaLI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2128)..(2133)
; OTHER INFORMATION: PstI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2204)..(2209)
; OTHER INFORMATION: NcoI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2284)..(2289)
; OTHER INFORMATION: AvaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2294)..(2299)
; OTHER INFORMATION: AvaI, SmaI, and XmaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2551)..(2556)
; OTHER INFORMATION: ApaLI site
; OTHER INFORMATION: AvaI, SmaI, and XmaI site
US-10-939-107-60

Query Match 97.2%; Score 548; DB 24; Length 2773;
Best Local Similarity 100.0%; Pred. No. 1.4e-169;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CATGATGATAAACAATGATGGTGCTAATGTTGCTTCAACAACAATTCGTGTGAACGTG 60
|||

Db 5 CATGATGATAAACAATGATGCTGCTAATGTTGCTTCAACAACAANTCTGTTGAACGTGTG 64
Qy 61 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 120
Db 65 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 124
Qy 121 GCACCTGCAAAAAAACAACGCTTTTGACGGGGCCCATACATAGTACAACTCTAGCTTTC 180
Db 125 GCACCTGCAAAAAAACAACGCTTTTGACGGGGCCCATACATAGTACAACTCTAGCTTTC 184
Qy 181 GTAGACTATTTTACATAAATAGCTACACCGTTGTATAGCTTCCAAATACACTACACAC 240
Db 185 GTAGACTATTTTACATAAATAGCTACACCGTTGTATAGCTTCCAAATACACTACACAC 244
Qy 241 AFTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGCAGTCACTAGGCGGGCCTTATC 300
Db 245 AFTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGCAGTCACTAGGCGGGCCTTATC 304
Qy 301 GGGTCGGCTGCTGACGTACGAATCACATTCGACCGGACGAGTGTGTCTTATCGT 360
Db 305 GGGTCGGCTGCTGACGTACGAATCACATTCGACCGGACGAGTGTGTCTTATCGT 364
Qy 361 GACAGCGCGCAGCTTCTGCTGCTAACCGCAGCGCGCACTCTTATCGGAACA 420
Db 365 GACAGCGCGCAGCTTCTGCTGCTAACCGCAGCGCGCACTCTTATCGGAACA 424
Qy 421 GACAGCGCGCTCCATATCAGCGCGCGGTATCTCATGCGCGTACCGGACACGAGCGGCC 480
Db 425 GACAGCGCGCTCCATATCAGCGCGCGGTATCTCATGCGCGTACCGGACACGAGCGGCC 484
Qy 481 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTTGAAACAG 540
Db 485 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTTGAAACAG 544
Qy 541 CATCTGTT 548
Db 545 CATCTGTT 552

RESULT 6
US-10-622-088-89
; Sequence 89, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth
; APPLICANT: Franke, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622,088
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/32437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 89
; LENGTH: 5038
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: pIB/V5-His-DEST
US-10-622-088-89
Query Match 96.9%; Score 546.4; DB 21; Length 5038;
Best Local Similarity 99.8%; Pred. No. 6.3e-169;
Matches 547; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CATGATGATAAACAATGATGCTGCTAATGTTGCTTCAACAACAANTCTGTTGAACGTGTG 60
Db 1 CATGATGATAAACAATGATGCTGCTAATGTTGCTTCAACAACAANTCTGTTGAACGTGTG 60
Qy 61 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 120
Db 61 TTTTCATGTTTCCCAACAGCACCTTTTATACCTCGGTGGCTCCCAACCACTTTT 120
Qy 121 GCACCTGCAAAAAAACAACGCTTTTGACGGGGCCCATACATAGTACAACTCTAGCTTTC 180
Db 121 GCACCTGCAAAAAAACAACGCTTTTGACGGGGCCCATACATAGTACAACTCTAGCTTTC 180
Qy 181 GTAGACTATTTTACATAAATAGCTACACCGTTGTATAGCTTCCAAATACACTACACAC 240
Db 181 GTAGACTATTTTACATAAATAGCTACACCGTTGTATAGCTTCCAAATACACTACACAC 240
Qy 241 AFTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGCAGTCACTAGGCGGGCCTTATC 300
Db 241 AFTGAACCTTTTTCAGTGCAAAAAAGTACGTGCGCAGTCACTAGGCGGGCCTTATC 300
Qy 301 GGGTCGGCTGCTGACGTACGAATCACATTCGACCGGACGAGTGTGTCTTATCGT 360
Db 301 GGGTCGGCTGCTGACGTACGAATCACATTCGACCGGACGAGTGTGTCTTATCGT 360
Qy 361 GACAGCGCGCAGCTTCTGCTGCTAACCGCAGCGCGCACTCTTATCGGAACA 420
Db 361 GACAGCGCGCAGCTTCTGCTGCTAACCGCAGCGCGCACTCTTATCGGAACA 420
Qy 421 GACAGCGCGCTCCATATCAGCGCGCGGTATCTCATGCGCGTACCGGACACGAGCGGCC 480
Db 421 GACAGCGCGCTCCATATCAGCGCGCGGTATCTCATGCGCGTACCGGACACGAGCGGCC 480
Qy 481 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTTGAAACAG 540
Db 481 GTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAACACAGTTGAAACAG 540
Qy 541 CATCTGTT 548
Db 541 CATCTGTT 548
RESULT 7
US-09-896-888A-14
; Sequence 14, Application US/09896888A
; Patent No. US20020116723A1
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Insect Expression Vectors
; FILE REFERENCE: 80021-44
; CURRENT APPLICATION NUMBER: US/09/896,888A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US/09/048,911
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/049,946
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Promoter
; OTHER INFORMATION: sequence of the OpNPV iel2 gene
US-09-896-888A-14

```
Query Match      81.9%; Score 462; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.5e-141; Indels 0; Gaps 0;
Matches 462; Conservative 0; Mismatches 0;

Qy 102 CCCCAACCACTCTAGTTTCGTAAGACTATTTTACATAAATAGTCTACACCGTTGATACGC 221
Db 1 CCCCAACCACTCTAGTTTCGTAAGACTATTTTACATAAATAGTCTACACCGTTGATACGC 60

Qy 162 AGTACAAACTCTAGTTTCGTAAGACTATTTTACATAAATAGTCTACACCGTTGATACGC 221
Db 61 AGTACAAACTCTAGTTTCGTAAGACTATTTTACATAAATAGTCTACACCGTTGATACGC 120

Qy 222 TCCAAATACACTACACACATTTGAACCTTTTTCAGTGCACAAAGTACGTCGCGAGT 281
Db 121 TCCAAATACACTACACACATTTGAACCTTTTTCAGTGCACAAAGTACGTCGCGAGT 180

Qy 282 CACGTAGCGCGCCCTTATCGGGTTCGCGTCTTCATACGTACGAATCACATTTATCGGACCGG 341
Db 181 CACGTAGCGCGCCCTTATCGGGTTCGCGTCTTCATACGTACGAATCACATTTATCGGACCGG 240

Qy 342 ACGAGTGTTCCTATTCGTGACAGACGCCAGCTTCCTGTGTTGCTAAACCGACCGGAC 401
Db 241 ACGAGTGTTCCTATTCGTGACAGACGCCAGCTTCCTGTGTTGCTAAACCGACCGGAC 300

Qy 402 GCAACTCCTTATCGAAGCAGGACGCGCTCCATATCAGCGCGCTTATCTCATGCGCGT 461
Db 301 GCAACTCCTTATCGAAGCAGGACGCGCTCCATATCAGCGCGCTTATCTCATGCGCGT 360

Qy 462 GACCGGACACAGGCGCCGTCGCGCTTATCGCGCTTATAATACAGCCGCAACGATCT 521
Db 361 GACCGGACACAGGCGCCGTCGCGCTTATCGCGCTTATAATACAGCCGCAACGATCT 420

Qy 522 GGTAAACACAGTTGAACAGCATCTGTTACAGCGACACAACT 563
Db 421 GGTAAACACAGTTGAACAGCATCTGTTACAGCGACACAACT 462
```

```
RESULT 8
; Sequence 127, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622,088
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/22437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 127
; LENGTH: 147
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombination region of pIB/V5-His-DEST
```

```
; NAME/KEY: misc feature
; LOCATION: (141)..(148)
; OTHER INFORMATION: n may be any nucleotide
US-10-622-088-127

Query Match      11.0%; Score 62; DB 21; Length 147;
Best Local Similarity 100.0%; Pred. No. 1.1e-09; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0;

Qy 487 CTTATCGCGCTATAAATACAGCCGCAACGATCTGTTAAACACAGTTGAACGATCTG 546
Db 1 CTTATCGCGCTATAAATACAGCCGCAACGATCTGTTAAACACAGTTGAACGATCTG 60

Qy 547 TT 548
Db 61 TT 62

RESULT 9
US-10-622-088-149
; Sequence 149, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622,088
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/22437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 149
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Recombination region of pIB/V5 His DEST
; NAME/KEY: misc feature
; LOCATION: (141)..(142)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (145)..(276)
US-10-622-088-149

Query Match      11.0%; Score 62; DB 21; Length 325;
Best Local Similarity 100.0%; Pred. No. 1.7e-09; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0;

Qy 487 CTTATCGCGCTATAAATACAGCCGCAACGATCTGTTAAACACAGTTGAACGATCTG 546
Db 1 CTTATCGCGCTATAAATACAGCCGCAACGATCTGTTAAACACAGTTGAACGATCTG 60

Qy 547 TT 548
Db 61 TT 62
```


RESULT 10

US-09-896-888A-16
; Sequence 16, Application US/09896888A
; Patent No. US20020116723A1
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Insect Expression Vectors
; FILE REFERENCE: 80021-44
; CURRENT APPLICATION NUMBER: US/09/896,888A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US/09/048,911
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/049,946
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 92
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fragment of
; OTHER INFORMATION: promoter sequence of the AcMNPV ien gene
US-09-896-888A-16

Query Match 10.1%; Score 57; DB 9; Length 92;
Best Local Similarity 77.5%; Pred. No. 4e-08;
Matches 69; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 186 CTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACACACATTGA 245
Db 1 CTATTTTATTAGTAATAGTCTACACTGTACGATACGCTCCCAATATACTACTACATATCA 60
QY 246 ACCTTTTTCAGTCGCAAAAAGTACGTTG 274
Db 61 ACTTTTTCGATTACAAAAAAGTTCAATT 89

RESULT 11

US-09-896-888A-15
; Sequence 15, Application US/09896888A
; Patent No. US20020116723A1
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Insect Expression Vectors
; FILE REFERENCE: 80021-44
; CURRENT APPLICATION NUMBER: US/09/896,888A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US/09/048,911
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/049,946
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 88
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Fragment of
; OTHER INFORMATION: promoter sequence of the AcMNPV ien gene
US-09-896-888A-15

Query Match 7.6%; Score 42.6; DB 9; Length 88;
Best Local Similarity 74.0%; Pred. No. 0.0022;
Matches 54; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 112 AACTTTTTCAGTCAAAAAACACGCTTTTTCACGCGGGCCCATACATAGTACAAACT 171
Db 16 AATTTTTCGATGCAAAAAAGTTCACTTTTCCTGACACTCCATATACAGTCAATCT 75

QY 172 CTACGTTTCGTAG 184
Db 76 CTACAAATCGTAG 88

RESULT 12

US-10-473-126-199/c
; Sequence 199, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 199
; LENGTH: 6289
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
US-10-473-126-199

Query Match 7.0%; Score 39.6; DB 21; Length 6289;
Best Local Similarity 53.2%; Pred. No. 0.18;
Matches 84; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 112 AACCTTTTTCAGTCTCAAAAAACACGCTTTTTCACGCGGGCCCATACATAGTACAAACT 171
Db 4896 AATTTATTATATACCAAAATAAAACTATTATATAATCCACTACCAAAATATCATTCAAAA 4837
QY 172 CTACGTTTCGTAGACTATTTTACATAAAATAGTCTACACGTTGTATACGCTCCAAATACA 231
Db 4836 TAATAATTTTAAACATTTTATATAAAATCGTCAAAATCTTTTACTCAAAAAATA 4777
QY 232 CTACACACATTGAACCTTTTTCGAGTGCAAAAAAGTA 269
Db 4776 ATACAACTTTCCTCTCCCTATAAAAAACGAAAAAATA 4739

RESULT 13

US-10-221-714A-223/c
; Sequence 223, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 223
; LENGTH: 9289
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:40:53 ; Search time 1072.29 Seconds
(without alignments)
7998.346 Million cell updates/sec

Title: US-09-896-888A-1_COPY_351_527
Perfect score: 177
Sequence: 1 gtcctatcgtgacagcagc.....gcccgcaacatctggtgtaa 177

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1_0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	177	100.0	564	6	BD070856		BD070856 Insect ex
C	2	177	100.0	1429	14	S64501	S64501 p8.9=8.9 kd
	3	177	100.0	2773	6	AX766573	AX766573 Sequence
	4	177	100.0	4170	14	NPHTTAA	M83827 Orgyia pseu
	5	177	100.0	131995	14	OPU75930	U75930 Orgyia pseu
C	6	35.2	19.9	129609	14	AF512031	AF512031 Choriston
C	7	34.8	19.7	300363	1	AE016781	AE016781 Pseudomon
	8	33	18.6	155274	10	AC121979	AC121979 Mus muscu
C	9	32.8	18.5	179571	2	CR361569	CR361569 Danio rer
	10	32.4	18.3	177263	8	AP003376	AP003376 Oryza sat
C	11	32.4	18.3	222856	2	AC134482	AC134482 Rattus no
	12	32.4	18.3	270958	2	AC135655	AC135655 Rattus no
C	13	32.2	18.2	299550	1	AP005031	AP005031 Streptomy
C	14	32	18.1	921	6	BD221352	BD221352 Human gen
C	15	32	18.1	131158	14	AV327402	AV327402 Choriston
C	16	31	17.5	110000	1	CP000010.02	Continuation (3 of
C	17	31	17.5	303450	1	SC0939130	AL939130 Streptomy
	18	30.8	17.4	1268	6	CQ597684	CQ597684 Sequence
	19	30.8	17.4	1342	6	CQ610737	CQ610737 Sequence

20	30.8	17.4	1426	3	AY070687	Drosophil
C 21	30.8	17.4	3448	6	CQ610736	Sequence
C 22	30.8	17.4	3600	6	CQ597683	Sequence
C 23	30.8	17.4	10172	1	AE012382	Xanthomon
C 24	30.8	17.4	17997	2	AC019883	Drosophil
C 25	30.8	17.4	36800	9	HS313D11	Human DNA s
C 26	30.8	17.4	41842	2	AC133915	Homo sapi
C 27	30.8	17.4	93375	3	AC005554	Drosophil
C 28	30.8	17.4	140403	2	AC109444	Homo sapi
C 29	30.8	17.4	164920	3	AC092187	Drosophil
C 30	30.8	17.4	184554	3	AC099022	Drosophil
C 31	30.8	17.4	256073	9	AE006464	Homo sapi
C 32	30.8	17.4	314957	3	AE003581	Drosophil
C 33	30.6	17.3	3000	9	AF331832S1	Homo sapi
C 34	30.6	17.3	110000	2	AC008576_0	Homo sapi
C 35	30.6	17.3	125020	9	AF429315	Homo sapi
C 36	30.6	17.3	189509	9	AC099564	Homo sapi
C 37	30.4	17.2	41905	9	AC142495	Homo sapi
C 38	30.4	17.2	105642	9	AL732414	Human DNA
C 39	30.4	17.2	119525	9	AC142496	Homo sapi
C 40	30.4	17.2	147127	9	AC142497	Homo sapi
C 41	30.2	17.1	1127	8	AK104298	Oryza sat
C 42	30.2	17.1	1128	8	AK070153	Oryza sat
C 43	30.2	17.1	2021	8	AK072256	Oryza sat
C 44	30.2	17.1	2127	8	AK107045	Oryza sat
C 45	30.2	17.1	33087	2	AC149421	Phakopsor

ALIGNMENTS

RESULT 1
BD070856
LOCUS BD070856 564 bp DNA linear PAT 27-AUG-2002
DEFINITION Insect expression vectors.
ACCESSION BD070856
VERSION BD070856.1 GI:22616459
KEYWORDS JP 2001516225-A/1.
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 564)
AUTHORS Grigliatti,T.A., Theilmann,D.A., Pfeifer,T.A. and Hegedus,D.D.
TITLE Insect expression vectors
JOURNAL Patent: JP 2001516225-A 1 25-SEP-2001;
THE UNIVERSITY OF BRITISH COLUMBIA
COMMENT OS Multicapsid nucleopolyhedrovirus
PN JP 2001516225-A/1
PD 25-SEP-2001
PF 26-MAR-1998 JP 1998541010
PR 27-MAR-1997 US 60/049946,28-JAN-1998 CA 2221819 PI
THOMAS A GRIGLIATTI,DAVE A THEILMANN,THOMAS
A PFEIFER,DWAYNE D
PI HEGEDUS
PC C12N15/85,C12N5/06,C12N15/69//C12N9/22
CC Insect expression vectors
FH Key Location/Qualifiers
FT source 1..564
FT /organism='Multicapsid nucleopolyhedrovirus'.
FEATURES
source
Location/Qualifiers
1..564
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'

Query Match 100.0%; Score 177; DB 6; Length 564;
Best Local Similarity 100.0%; Pred.No. 6.2e-41;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCTATCGTGACAGCAGCCAGCTTCCTGCTGCTTAACCGCAGCGCAACTCCT 60
DB 351 GTCCTATCGTGACAGCAGCCAGCTTCCTGCTGCTTAACCGCAGCGCAACTCCT 410

Qy 61 TATCGGAACAGGCGCGCTCATATACAGCGCGGTTATCTCATGCGCGTACCGGACA 120
Db 411 TATCGGAACAGGCGCGCTCATATACAGCGCGGTTATCTCATGCGCGTACCGGACA 470
Qy 121 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 177
Db 471 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 527
RESULT 2
S64501/c
LOCUS S64501 1429 bp DNA linear VRL 30-SEP-1993
DEFINITION Orgyia pseudotsugata multicapsid nucleopolyhedrovirus
ACCESSION S64501
VERSION S64501.1 GI:404518
KEYWORDS
SOURCE
ORGANISM Orgyia pseudotsugata multicapsid nucleopolyhedrovirus
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE
AUTHORS Wu,X., Stewart,S. and Theilmann,D.A.
TITLE Characterization of an early gene coding for a highly basic 8.9K protein from the Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus
JOURNAL J. Gen. Virol. 74 (Pt 8), 1591-1598 (1993)
MEDLINE 93346965
PUBMED 8345350
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gisseq 136152] from the original journal article.
FEATURES
source
1..1429
/organism="Orgyia pseudotsugata multicapsid nucleopolyhedrovirus"
/mol_type="genomic DNA"
/db_xref="taxon:164623"
763..990
/gene="p8.9"
/note="8.9 kda basic protein"
763..990
/gene="p8.9"
/note="8.9 kda basic protein"
/codon_start=1
/protein_id="AAB27738.1"
/db_xref="GI:404519"
/translation="MNSQKIRMAKQQQVVARQHRRAAKLGRLYKAKKLRAELCEKLQ LQRVNDAALAKAFEEBFVYPHFSFYLTIN"
ORIGIN
Query Match 100.0%; Score 177; DB 14; Length 1429;
Best Local Similarity 100.0%; Pred. No. 6e-41;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTCCTATCGTGACAGGAGCCAGCTTCCTGTGTGCTAAACGCGAGCGGCAACTCCT 60
Db 415 GTCCTATCGTGACAGGAGCCAGCTTCCTGTGTGCTAAACGCGAGCGGCAACTCCT 356
Qy 61 TATCGGAACAGGCGCGCTCATATACAGCGCGGTTATCTCATGCGCGTACCGGACA 120
Db 355 TATCGGAACAGGCGCGCTCATATACAGCGCGGTTATCTCATGCGCGTACCGGACA 296
Qy 121 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 177
Db 295 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 239
RESULT 3
AX766573
LOCUS AX766573 2773 bp DNA linear PAT 25-JUN-2003
DEFINITION Sequence 60 from Patent WO03042244.
ACCESSION AX766573

VERSION AX766573.1 GI:32260450
KEYWORDS
SOURCE
ORGANISM synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1
AUTHORS Klynsner,S., Nielsen,F.S., Bratt,T., Voldborg,B. and Mouritsen,S.
TITLE Novel immunogenic mimetics of multimer proteins
JOURNAL Patent: WO 03042244-A 60 22-MAY-2003;
Pharmexa A/S (DK) ; Klynsner, Steen (DK) ; Nielsen, Finn Stausholm (DK) ; Bratt, Tomas (DK) ; Voldborg, Bjorn (DK) ; Mouritsen, Soren (DK)
FEATURES
Location/Qualifiers
1..2773
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="p22Op2F expression vector for insect cells"
561..566
/note="HindIII site"
573..578
/note="Aval site"
586..591
/note="EcoRI site"
593..598
/note="BamHI site"
625..630
/note="ClaI site"
629..634
/note="ClaI site"
1156..1161
/note="ApaLI site"
2128..2133
/note="PstI site"
2204..2209
/note="NcoI site"
2284..2289
/note="Aval site"
2294..2299
/note="Aval, SmaI, and XmaI site"
2551..2556
/note="ApaLI site"
ORIGIN
Query Match 100.0%; Score 177; DB 6; Length 2773;
Best Local Similarity 100.0%; Pred. No. 5.9e-41;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTCCTATCGTGACAGGAGCCAGCTTCCTGTGTGCTAAACGCGAGCGGCAACTCCT 60
Db 355 GTCCTATCGTGACAGGAGCCAGCTTCCTGTGTGCTAAACGCGAGCGGCAACTCCT 414
Qy 61 TATCGGAACAGGAGCGCGCTTCATATACAGCGCGGTTATCTCATGCGCGTACCGGACA 120
Db 415 TATCGGAACAGGAGCGCGCTTCATATACAGCGCGGTTATCTCATGCGCGTACCGGACA 474
Qy 121 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 177
Db 475 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 531
RESULT 4
NPHTTAA
LOCUS NPHTTAA 4170 bp DNA linear VRL 02-AUG-1993
DEFINITION Orgyia pseudotsugata nuclear polyhedrovirus transcriptional trans-activator (IE-2) gene, complete cds; ORF, complete cds.
ACCESSION M83827
VERSION M83827.1 GI:332540
KEYWORDS transcriptional transactivator.
SOURCE Orgyia pseudotsugata single capsid nucleopolyhedrovirus
ORGANISM Orgyia pseudotsugata single capsid nucleopolyhedrovirus Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.

```

REFERENCE 1 (bases 1 to 4170)
AUTHORS Theilmann,D.A. and Stewart,S.
TITLE Molecular analysis of the trans-activating IE-2 gene of Orgyia
pseudotsugata multicapsid nuclear polyhedrosis virus
JOURNAL Virology 187 (1), 84-96 (1992)
MEDLINE 92142536
PUBMED 1736546
REFERENCE 2 (bases 1 to 4170)
AUTHORS Theilmann,D.A. and Stewart,S.
TITLE Tandemly repeated sequence at the 3' end of the IE-2 gene of the
baculovirus Orgyia pseudotsugata multicapsid nuclear polyhedrosis
virus is an enhancer element
JOURNAL Virology 187 (1), 97-106 (1992)
MEDLINE 92142537
PUBMED 1736547
COMMENT Original source text: Orgyia pseudotsugata nuclear polyhedrosis
virus DNA.
FEATURES
    source
        1..4170
        /organism="Orgyia pseudotsugata single capsid
        nucleopolyhedrovirus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10450"
    gene
        627..3959
        /gene="IE-2"
    TATA_signal
        627..634
        /gene="IE-2"
    mRNA
        657..>1908
        /gene="IE-2"
        691..1908
        /gene="IE-2"
    CDS
        /codon_start=1
        /product="transcriptional trans-activator"
        /protein_id="AAA46749.1"
        /db_xref="GI:332541"
        /translation="MSRNNANAPTPSRRNLSLVGRRLTYSPDASQRAQSPRR
        SAPRAARVHAGDPGLPASYALPNGLHGDHFNPEEDDDILFVDTAAEQ
        AQRVNLHEAVNRHRELRGEMTRSPPTLNYSYSPSPRSRSPDLIMPEDLP
        AREQPLAFNDDDDLEQLLEQVLESAEPQLPQAPPAQVDVSLCHICSTFDIQ
        NNSNFTSTECNHAFCVKYVSI VFGKESYKSCINRTTISCRAYNRDGVSLTMS
        TVRDSQALKRWHAQLSDSNMPSHNMWTTIQELQAELELRARAHHDVNMARSDSO
        LLRQQLDVKEALEAHESVARLKLQKQNETLSNLSLQHLNTQVIESRVKMEQKQR
        HDEFWEKFLSLUS"
        1958..1963
        /gene="IE-2"
        2017..2759
        /gene="IE-2"
        2915..2922
        /gene="IE-2"
        2944..>3907
        /gene="IE-2"
        2984..3907
        /gene="IE-2"
        /note="ORF; homologous to the ACMPV PE-38 gene; putative"
        /codon_start=1
        /protein_id="AAA46750.1"
        /db_xref="GI:332542"
        /translation="WSSRYSPVIRNVGERRRVOBRLLSEFNSRPTVATAVACVLE
        TYCVQSNLIIDLPSECTHFCYKCVLNMVKNVPRAAVSCPMCKKVTQWQAF
        PNVSCFPIKTGTQRTFACIQFMALKTIQDRYNATEBEAEFSPVKMLQQLDA
        AOKEARDOENRQKQAHVNWNSCQVLTALQTLADMQQLRSEALSLEAHN
        RAANVQIDSLRAVLEAAQSPVSVNVFEFNDNARQNTLHFRPSRYVSTVSDMI
        EDSIKLSQSHVFGAACLPCSVNVVEINFPEDE"
        3954..3959
        /gene="IE-2"
    polyA_signal
    ORIGIN
    Query Match 100.0%; Score 177; DB 14; Length 4170;
    Best Local Similarity 100.0%; Pred. No. 5.8e-41;
    Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    1 GTCCTATCGTGACAGCGCCAGCTTCCTGTTGCTTAACCGCAGCGGACCGCAACTCCT 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Db 480 GTCCTATCGTGACAGCGCCAGCTTCCTGTTGCTTAACCGCAGCGGACCGCAACTCCT 539
Qy 61 TATCGGAACAGGACGCGCTCCATATCAGCGCGGCTTATCTCATCGCGTGCACCGGACA 120
|||||
Db 540 TATCGGAACAGGACGCGCTCCATATCAGCGCGGCTTATCTCATCGCGTGCACCGGACA 599
|||||
Qy 121 CGAGGCGCCGTCGCGCTTATCGGCGCTTAAATACAGCCGCGACGATCTGGTAAA 177
|||||
Db 600 CGAGGCGCCGTCGCGCTTATCGGCGCTTAAATACAGCCGCGACGATCTGGTAAA 656
|||||
RESULT 5
OPU75930 131995 bp DNA circular VRL 25-MAR-2003
LOCUS Orgyia pseudotsugata multicapsid nucleopolyhedrovirus, complete
genome.
U75930
ACCESSION U75930.2 GI:11024985
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 131995)
AUTHORS Ahrens,C.H., Russell,R.L., Funk,C.J., Evans,J.T., Harwood,S.H. and
Rohrmann,G.F.
TITLE The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
polyhedrosis virus genome
JOURNAL Virology 229 (2), 381-399 (1997)
MEDLINE 97271300
PUBMED 9126251
REFERENCE 2 (bases 1 to 131995)
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1996) Oregon State University, Agricultural
Chemistry, Corvallis, OR 97331-7301, USA
REFERENCE 3 (bases 1 to 131995)
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (06-MAR-1998) Oregon State University, Agricultural
Chemistry, Corvallis, OR 97331-7301, USA
REMARK Sequence update by submitter
REFERENCE 4 (bases 1 to 131995)
AUTHORS Rohrmann,G.F.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2000) Oregon State University, Agricultural
Chemistry, Corvallis, OR 97331-7301, USA
REMARK Sequence update by submitter
COMMENT On Oct 26, 2000 this sequence version replaced gi:2934903.
FEATURES
    Location/Qualifiers
        1..131995
        /organism="Orgyia pseudotsugata multicapsid
        nucleopolyhedrovirus"
        /mol_type="genomic DNA"
        /db_xref="taxon:164623"
        /note="similar to Autographa californica nuclear
        polyhedrosis virus (ACMPV) deposited in GenBank Accession
        Number L22858"
        complement(123..947)
        /note="ORF1; pk-1; similar to ACMPV ORF10"
        /codon_start=1
        /product="protein kinase 1"
        /protein_id="AAC59000.1"
        /db_xref="GI:1911247"
        /translation="MDAALQSLRFDVADCAVLAPKVVNGRCKMDVLRPTTSKLF
        RKTAAISFSADEINVDLMDSHSFVDMYFCYSPTAWAIVMDYVPCDLEETLQ
        GALDNALVNIVRQLCDALNDLNATGVIHNDVLEKLVLYFGARVLYLQGLCKRE
        HSPVHDTGLEYSPEKRRHRYARSFDWYAVGLAYKLTGCRHFFERSADVELSLAS
        NRRRQXNDPAALKNVRNLNMRADFVCLTFRENFECRSTDYKQIAKHSFLASHDYI"
        946..2370
        /note="ORF2; similar to ACMPV ORF9"
        /codon_start=1
        /product="1629-capsid"
    CDS

```

/protein_id="AAC59001.1"
/db_xref="GI:2934904"
/translation="MMERQYQSVRSYSLINQHNHAIAGPFLORVAGPEAHSVGRNVGD
RAVLNQAVLDLKLAEIDYADTAYMQAOPEASSRHFATLNRMLULLLIGVQDPDAR
RLNSVLARLEALLRQVNDVDAENVNLGDFEYSEYKISYQQTFAQTPTASASQQTQ
RLSPRQTSPLPQTQASLPQTPDQPEMWSPSPSVFHTTFAILLPQTTPATDTFSRP
SDEYVYVGKERAVDFRFPVPKPEHLKSRSSVATNAAGATPVAPPPPPPSAD
VTTSMPPPPPPPSADVTTSMPPPPPPVMDLATSMPPPPPPPPMVMDLATSMPPPINNA
INILLIDAMVAETNKAGDNRSALLDQIKQKTLKKTQPADGAPATPRTSLISEIRQ
GKTLKKRKEDQSGSTQTLKVDVDTTKTLKFNVTNIDRISKQEQEBKRDLDIT
KRPAVBDGNSNGNSDDWRDD"
complement (2533. .3270)
/note="ORF3; ph; similar to AcMNPV ORF8"
/codon_start=1
/product="polyhedrin"
/protein_id="AAC59002.1"
/db_xref="GI:1911249"
/translation="MPDYSRPTTGRYYVDNKKYKSLGSIKNAKKKHLLEHEEDE
KHLDPDHYMVAEDPFLGPGNQKLTFLKEIRNVKPDTKMLIWNMSKEFLRETWTFRF
VEDSPPIYNDQENVDFLVNMRPTRNRCYKFLAQHALWDCDVPHEVIRIVERPV
VGMNNEVRLISLAKGGCPTMNIHAETNSFESFVNRVIMENFKPIVYIGTDSSEE
ELLIELSVLKVKEFAPDAELFTGPAY"
complement (3350. .3541)
/note="ORF4"
/codon_start=1
/product="unknown"
/protein_id="AAC59003.1"
/db_xref="GI:1911250"
/translation="MRKNNGASQYPEHLTTVHKHRLQLSCVWCDDQRSDPHSAKGL
MFEHMFCLRIYVQQIKH"
complement (3698. .4093)
/note="ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAC59004.1"
/db_xref="GI:1911251"
/translation="MGASQSEQIKLKYRWSQTAALDDEKDLHLCLYDLERFVGA
HLNKRDQKARKKCAEAKALRVIEAADRHMLEAAAPACADGDRWSTLSRAQLDDI
AREKEIVDRHRLQKDSLLKQDRLLKR"
complement (4125. .4739)
/note="ORF6; left-2; similar to AcMNPV ORF6"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAC59005.1"
/db_xref="GI:1911252"
/translation="MERWNPAAIGDGLKRSETVLDVDPHDFVGVLTLSPVTVFERGLP
VRMSGMLLALLAPKQEPQOPAPRFPQSRNRNVCLKACDAQSLAKVLAARVMP
PCKMTWADLSSAPGRNRYKRFENCYLANVITCTKCTACLGLLHFYRMDAKCV
GEVTHLLIKAQDVYKPSNCAKMKVTKLCPQASCKGLNPTCNF"
complement (4742. .4975)
/note="ORF7; similar to AcMNPV ORF5"
/codon_start=1
/product="unknown"
/protein_id="AAC59006.1"
/db_xref="GI:1911253"
/translation="MNRPTMRNTAAVTTDYDREQLRRELNSLRSSVHELCTRSATGPD
CNRELEAGDRAPAVIVKAAAGGQSHSLCDKV"
5020. .5460
/note="ORF8; similar to AcMNPV ORF4"
/codon_start=1
/product="unknown"
/protein_id="AAC59007.1"
/db_xref="GI:1911254"
/translation="MPGARFVRFSLRTOBFKENVAVHMLGLRALIDGKVTSDAV
RRFGLGRNLVLSACMAVNVQYAPDATIONRHQPTIYFRVCQCHAMADVPPAPDDHS
IARYLLAECGAVLVDPDVFGEETEGVELLEVQRINAGGDI"
complement (5669. .6151)
/note="ORF9; ptp-2"
/codon_start=1
/product="protein tyrosine phosphatase 2"
/protein_id="AAC59008.1"
/db_xref="GI:1911255"
/translation="MYEASQIDEHFVFGYGYGDNEAMLRFTIENHAIASVISLSDSVA

PIRHALGLPYGDHIIHVYCEAAPCAALPNAMPALYDVMRRRIGIGEKRVLVHVCYAGASR
"SALAAVYVLMRSQMPYQDALNAVQSKRRVAINDHFVFLATRCYSYRFVNDLKPQIV
complement (6129. .6791)
/note="ORF10; ptp-1; similar to AcMNPV ORF1"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAC59009.1"
/db_xref="GI:1911256"
/translation="MLTSRRWAVIYTRTSRNCGGAWCTNGVRRRRQVHLPSPVARAM
FPRMHEYTAGAVIEGTRLLCFKPLNAELFEYVSDERWTNAASVLAHSALGAVI
DLTNTARYDGAQVMKGLYKKIRVPEGRAVPDDDI VAEFIETVEEFRCPTMLVAV
HWTHGLNRSGLVCRYMVERLGSPTDAIRFETARGHKIERTNYLOLLARKHVRGQ
PN"
6819. .7814
/note="ORF11; similar to AcMNPV ORF11"
/codon_start=1
/product="unknown"
/protein_id="AAC59010.1"
/db_xref="GI:1911257"
/translation="MSLSKLLVAVYGSYNLPHDRYGESVHLVRIVHEHLTNTVVS
ASCVRDIATARCLNSGHLCFEDVARQLLDNSEVAARLSANFRCGDATGLCLDMORALIA
DIDRHAPLARVGRRANIFALDAIADIPSDVNNLQIIGRFMHFFPCSGLARVAVF
DPDIRADGWYHKFCVLTVMHLVACGAPAGSATRLRDADAVKHGHPNDEGNCAPAI
VYGRFCAIGREHFAHKTACMHLIFQFMRNDLTPADERHPCFVGKIDGQCKDITYD
LRTHADALYTHGTTDROKNAFLDLLCCVNASDIDADCYDCVNVKFYATQNKYKWM"
complement (8180. .9142)
/note="ORF12; similar to AcMNPV ORF13"
/codon_start=1
/product="unknown"
/protein_id="AAC59011.1"
/db_xref="GI:1911258"
/translation="MFSWMEGWMNADEOVNABEEOAYRYAVDQFAHSDLVBRDVF
RCHPFFVKRYVLDETAGRCSSVDFCKGLKISHDLLQRCNDRQHVRLQNLVILGAP
PAQPDLSGLSFATKGLVQLLQFSPCKNEVLAVGANKDHFDRQHVRLQNLVILGAP
TLNTNSDKFIISAKSEFLVGARFEQRLQTLTKNALQCAAPRTAPGVVFRP
DVTKHPLHAFVGRVEDRGVTQIFARGQEHRFKRLEFEGMDVDVGRGNPLLA
VHCICEFANGHKIRLPKKVTEVCAVNVAKDIVKAILNKT"
complement (9094. .9825)
/note="ORF13; left-1; similar to AcMNPV ORF14"
/codon_start=1
/product="late expression factor 1"
/protein_id="AAC59012.1"
/db_xref="GI:1911259"
/translation="MAPCKYTPERVOMMDAIAYNDSRRILAFMTDRPRWVHAHNFDS
AAQLFAIVKNSISDVHVXKLEBEGREWVIDADPKCDAKAEMLKVNVGATAPMLPF
EGEDAVQITMFSNGRGFWHLWPCFKGKADAPKSLREHWNFVKPAKLVSGDIRPG
SFADCVRRVAVHYIGDAREDLVRYWPDVDRDVFNCANKOIRAFPSYNYKGGDYSRCL
TOOLQORIKACSAGCLAGGTPTSK"
9893. .11362
/note="ORF14; egt; similar to AcMNPV ORF15"
/codon_start=1
/product="ecdysteroid UDP-glucosyltransferase"
/protein_id="AAC59013.1"
/db_xref="GI:1911260"
/translation="MVFLIALLTLATGARAASILAVLTPPAYSHVVYRAYVHALVK
NCHNVTIKQLLDYAVQDECGVEQIDADMSAQYKKLVASSGVFKRGVVADETTV
TADNYMGLIEMFKDQPDNANVRFLSTNRTFDVAVVEAFADYALVFGHLFEPASVIOI
APGYLAENFERRAVRAPHLHYPTFGAALTIRGGALSERLNEFELIARRSDELL
KQFGKSTPTTIQRLDNVQLLLNHPVYDNNRPVPSVQYLGGLHLAQLPQLDA
PLERRLNSVDGAVYVSGSIDTNSIHAFLMLLDTFANLNNTYTLKVVDDAAS
VALPRNVLAQKWFSTAVLRKNVAFVTOAGLQSSDEALQARQVPMVCLPMMGQFHH
ARKLQOQGVARALDTAAVSAPOLQAIREVADGEAYRARIDKLRAVVEHDAAPDEKA
VKFTERVIKNNVNVNWFARSLKTTAANWASIDYFVRFL"
11547. .12140
/note="ORF15; similar to AcMNPV ORF16"
/codon_start=1

Query Match 100.0%; Score 177; DB 14; Length 131995;

Best Local Similarity 100.0%; Pred. No. 5.2e-41;

Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCTTATCGTACAGGACGCCAGCTTCCTGTGTTGCTAAACCGCAGCGGACGCAACTCCT 60

```

Db 128522 GTCCTATCGTACAGACGCCAGCTTCCTGTGTTGCTAACCGCAGCGGACCAACTCCT 128581
QY 61 TATCGGAACAGCAGCGCTCCATATCAGCCGCGCTTATCTCATGCGGTGACCGGACA 120
Db 128582 TATCGGAACAGCAGCGCTCCATATCAGCCGCGCTTATCTCATGCGGTGACCGGACA 128641
QY 121 CGAGCGCGCGTCCGCTTATCGCGCTATAAATACAGCCCAACGATCTGGTAAA 177
Db 128642 CGAGCGCGCGTCCGCTTATCGCGCTATAAATACAGCCCAACGATCTGGTAAA 128698

RESULT 6
AF512031/c 129609 bp DNA circular VRL 27-MAY-2004
LOCUS
DEFINITION
Choristoneura fumiferana MNPV polyhedrin, complete genome.
ACCESSION
AF512031 AF177329 S78506 S81690 U10441 U18677 U26676 U26734 U53854
U57401 U59008 U70432 U72240 X65395 S46001
VERSION
AF512031.2 GI:47157118
KEYWORDS
Choristoneura fumiferana MNPV
SOURCE
Choristoneura fumiferana MNPV
ORGANISM
Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
REFERENCE
1 (bases 1 to 129609)
Lee,H.Y., Arif,B., Dobos,P. and Krell,P.
Identification of bent DNA and ARS fragments in the genome of
Choristoneura fumiferana nuclear polyhedrosis virus
Virus Res. 24 (3), 249-264 (1992)
JOURNAL
93033705
MEDLINE
1413988
PUBMED
2 (bases 1 to 129609)
Xie,W.D., Arif,B., Dobos,P. and Krell,P.J.
Identification and analysis of a putative origin of DNA replication
in the Choristoneura fumiferana multinucleocapsid nuclear
polyhedrosis virus genome
Virology 209 (2), 409-419 (1995)
JOURNAL
95297142
MEDLINE
7778276
PUBMED
3 (bases 1 to 129609)
Liu,J.J. and Carstens,E.B.
Identification, localization, transcription, and sequence analysis
of the Choristoneura fumiferana nuclear polyhedrosis virus DNA
polymerase gene
Virology 209 (2), 538-549 (1995)
JOURNAL
95297155
MEDLINE
7778286
PUBMED
4 (bases 1 to 129609)
Barrett,J.W., Krell,P.J. and Arif,B.M.
Characterization, sequencing and phylogeny of the ecdysteroid
UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis
viruses isolated from Choristoneura fumiferana
J. Gen. Virol. 76 (Pt 10), 2447-2456 (1995)
JOURNAL
96030854
MEDLINE
7595348
PUBMED
5 (bases 1 to 129609)
Qiu,W., Liu,J.J. and Carstens,E.B.
Studies of Choristoneura fumiferana nuclear polyhedrosis virus gene
expression in insect cells
Virology 217 (2), 564-572 (1996)
JOURNAL
96183379
MEDLINE
8610448
PUBMED
6 (bases 1 to 129609)
Liu,J.J. and Carstens,E.B.
Identification, molecular cloning, and transcription analysis of
the Choristoneura fumiferana nuclear polyhedrosis virus
spindle-like protein gene
Virology 223 (2), 396-400 (1996)
JOURNAL
96400202
MEDLINE
8806578
PUBMED
7 (bases 1 to 129609)
Lapointe,R., Back,D.W., Ding,Q. and Carstens,E.B.
Identification and molecular characterization of the Choristoneura

```

```

fumerana multicapsid nucleopolyhedrovirus genomic region encoding
the regulatory genes pkip, p47, lef-12, and gta
Virology 271 (1), 109-121 (2000)
JOURNAL
20276145
MEDLINE
10814576
PUBMED
8 (bases 1 to 129609)
Carstens,E.B., Liu,J.J. and Dominy,C.
Identification and molecular characterization of the baculovirus
CFMNPV early genes: ie-1, ie-2 and ps38
Virus Res. 83 (1-2), 13-30 (2002)
JOURNAL
21854555
MEDLINE
11864738
PUBMED
9 (bases 1 to 129609)
de Jong,J.G., Dominy,C.N., Lauzon,H.A., Arif,B.M., Carstens,E.B.
and Krell,P.J.
Complete Genome of Choristoneura fumiferana Multiple
Nucleopolyhedrovirus
Unpublished
JOURNAL
10 (bases 1 to 129609)
de Jong,J.G., Dominy,C.N., Lauzon,H.A., Arif,B.M., Carstens,E.B.
and Krell,P.J.
Direct Submission
Submitted (13-MAY-2002) Department of Microbiology, University of
Guelph, Chemistry and Microbiology Building, Guelph, ON N1G 2W1,
Canada
JOURNAL
11 (bases 1 to 129609)
de Jong,J.G., Dominy,C.N., Lauzon,H.A., Arif,B.M., Carstens,E.B.
and Krell,P.J.
Direct Submission
Submitted (13-MAY-2004) Department of Microbiology, University of
Guelph, Chemistry and Microbiology Building, Guelph, ON N1G 2W1,
Canada
REMARK
Sequence update by submitter
COMMENT
On May 13, 2004 this sequence version replaced gi:30269978.
FEATURES
Location/Qualifiers
1..129609
/organism="Choristoneura fumiferana MNPV"
/mol_type="genomic DNA"
/db_xref="taxon:208973"
/country="Ireland"
/complement(1..735)
/contig="ORF1; Ac8/Op3"
/codon_start=1
/product="polyhedrin"
/protein_id="AAP29795.1"
/db_xref="GI:30269978"
/translation="MPDYSRPPTIGRTYVDNKYKNGLSVVKRKKHLLHEBDEK
HLDLDHYMVAEDFLGKQKLTFLKRIKRVKPDTHKLRWDCDYPVHEVIRIVEPSYV
EDSPFVINDQVMDVFLVNNRPTRNRCYKFLRQHLRWDYVPEVIRIVEPSYV
GMNNEYRISLAKGGCPINNIHAEYTNSEFVNRVWENFYKPIVYIGTDSGEE
MLIEVSLVFKVEFADPLPTGPAY"
/complement(838..1449)
/contig="ORF2; Op5"
/codon_start=1
/product="unknown"
/protein_id="AAP29796.1"
/db_xref="GI:30269980"
/translation="MQGVFDRRAALLQDCQMGDGFVDSREHIQAVQLRENE
GRKVPKSKHVOGLESNQLLVGVSKHGWAASDQKILELFYRNSSTGTGEMLDDEK
DLHLYDLERFVGAHLNRCDAKKNKCAKALKRVQISADRHAALEAAAAIPDADD
GQWSKLTSQQLDEIAREKDIIVRYQLQKQDRLIKMDGLKKQ"
/complement(1251..1859)
/contig="ORF3; lef2; Ac6/Op6"
/codon_start=1
/product="late expression factor 2"
/protein_id="AAP29797.1"
/db_xref="GI:30269981"
/translation="MDQVWNPAGAGSVKPEYTLIDPNDFVGMLELTPYTVFERGLF
IRMSGMLLALLAAPKPKVPTTRFPORSKRNVCLKECADGPOSILTQVLTARMNPLC
MSKIMADLGSAPRCNMVYKKEFNLCYLANVLTCTKCAACLIIGALLHLYRMDAKCVGE
VTHLLIKAENTYKESNCAMKAVTKLCPKANMCKGLNPICNY"
/complement(1862..2092)
/contig="ORF4; Ac5/Op7"

```

CDS /codon_start=1
/product="unknown"
/protein_id="AAP29798.1"
/db_xref="GI:30269982"
/translation="MTRPTMRNAAVAADYDREQLRRDLNSLRSSVHELCTRSTTGFD
CNRELDGVDKAPAVIIKPAAGQSHSLCDKV"
2137..2568
CDS /note="ORF5; Ac4/Op8"
/codon_start=1
/product="unknown"
/protein_id="AAP29799.1"
/db_xref="GI:30269983"
/translation="MPNFVKSRLTRSKENIVAHIGHLSRVRELIDRNVTPADVRR
FRFPDRALLAACVNVQAYGADGTIRLQPTLLYYVCRNFRVADVPAPDDHFIAR
YLWPECSTPLVHLPLHVFGDTEGVNELLEVQINAGDL"
2669..2860
CDS /note="ORF6; Ac2"
/codon_start=1
/product="baculovirus repeated ORF"
/protein_id="AAP29800.1"
/db_xref="GI:30269984"
/translation="MDNRNKKNAVTDWIFMNCIHYTRTGIIHHPCLSNSFVKRPVLYQ
PHTVLIINNTAQADAPQAI"
complement(2896..3699)
CDS /note="ORF7; Ac136/Op132"
/codon_start=1
/product="p26"
/protein_id="AAP29801.1"
/db_xref="GI:30269985"
/translation="MAMLTFFLLICVSSAALS VKVTTINNVLVTNDTKTIQVAKQVD
DKPAIVQIPQSFQKQELDLHHPGVASNVMPFRANNTKLTLLVLDGSLATIT
DVRYTTFHSHKMMVYQOLYSPALSNFSLANOIYIGAPIFEKERMVSVITAHEDYK
NKLVIYPTGISARGLSGQINFDLQTLTKLEGSSVYGMQLPYKALKDYALSTNR
KNJFKGLPRNAVYFNERDITIALVEGSEFIDRILRUSGPLILNKIKQQ"
complement(3738..4320)
CDS /note="ORF8; ptp2; Op9"
/codon_start=1
/product="protein tyrosine phosphatase 2"
/protein_id="AAP29802.1"
/db_xref="GI:30269986"
/translation="MYDANQIDKYIFVGGYGD DKAMLOFTKKYDIASVISLINADVG
PIRQALGPAGEHIVYCEADPTCVLPNMFALYETMRIRIGEGKRLVHCYAGESR
SAALVYVYMRSRQMSYEALS LKVKRRVAISNHFVRLASKCSYKFNANVLKIRVS
"
complement(4198..4731)
CDS /note="ORF9; ptp1; Ac1/Op10"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAP29803.1"
/db_xref="GI:30269987"
/translation="MFDRWHEFTPCGRVIDTGLICEFKYPLSAELFEYVNDNRWT
VASLTHRSALGAVIDITNTTRYDGMIREGLLYKKRVPGRRAIDEDTVQKFFSA
VDFQDCPTMLGVHCTHGLNRSGLVCRYMWDKLGVS PADAIRFEAEARGHKIERA
NYLQDLLARNHVRREP"
4800..5819
CDS /note="ORF10; Ac11/Op11"
/codon_start=1
/product="unknown"
/protein_id="AAP29804.1"
/db_xref="GI:30269988"
/translation="MHSVRNLFDSNMSLSKLLVYAYGANLLHEKYGSYHLRYTV
HEHTTVSNASCVRDITATRCFENGFCFNARQLLDVTDVAARLAAMVNRGDKTG
LCANVQALAEIDKXAPLEKVSIGNIFALDTIADIPSNALDPFOI IIEGPKDFVD
MNNJAHVADVPDPKIKAEQWYIKFCVLTIMHRLTVNAVPTBELMRLQDAVIAKFVQ
PQNGKCAPANAYPGCGIGRKHFSQHKAAASMYILFQYMRNNLLTPKDERHPSFGVI
KDFGLRCKEYTDLRAEADLYINATTDKKNALFDLLCCVNAADLDCVYDIVDNF
YFICK"
complement(5854..6387)

Query Match 19.9%; Score 35.2; DB 14; Length 129609;
Best Local Similarity 92.5%; Pred. No. 11;
Matches 37; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 137 CTTATCGCGCCTATAATACAGCCCGCAACGATCTGGTAA 176 linear BCT 12-DEC-2002
|||||
Db 124634 CTTATCGCGCCTATAATACAGCCCGCAACGATCTGGTAA 124595
|||||
RESULT 7
AE016781/c 300363 bp DNA linear BCT 12-DEC-2002
LOCUS Pseudomonas putida KT2440 section 8 of 21 of the complete genome.
DEFINITION AE016781 AE015451
ACCESSION AE016781.1 GI:26557025
VERSION
KEYWORDS
SOURCE Pseudomonas putida KT2440
ORGANISM Pseudomonas putida KT2440
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 300363)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holtzapple,E., Scanlan,D., Iran,K., Moazzes,A., Uterback,I.,
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Fraser,C.
Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440
Environ. Microbiol. 4 (12), 799-808 (2002)
2 (bases 1 to 300363)
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
Nelson,M., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
Holtzapple,E., Scanlan,D., Iran,K., Moazzes,A., Uterback,I.,
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
Duesterhoft,A., Tummeler,B. and Fraser,C.
Direct Submission
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..300363
/organism="Pseudomonas putida KT2440"
/mol_type="genomic DNA"
/strain="KT2440"
/db_xref="taxon:160488"
78..1985
/locus_tag="Pp1883"
78..1985
/locus_tag="Pp1883"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAN67502.1"
/db_xref="GI:24983391"
/translation="MLTPISFARCSGSAVLYLCOLSNRKCARRYSYVMDPVDYSHLN
ERNVVDLALGASADLMTWISYEGIAEHDLIYINRWCKKADNSREDLFGREVLSDLA
GRGEILDNALVTALRGQVYFSTRUDNAGNPVSGSKRRFFVDRPAGGIWALPVA
HFQDAHEGVLDVNLNPEGILLATAPYAAMAKDDKVELSWEPMLNESLPGVPFFQVLE
VQSDPGRLLMLHDIGQVFMYLGGFAFLHYRIFEADGTSRSQVRFDIYSPAAGDP
PPAPLLAASPIPGATDGLDPPDDEYTRAGLLEVDVPELGLDGLVLYAVGPVTLR
SLAVDASVDSKRLTHELEREMQSEVNRGKRFAFSVEFSRQASQVLDLVLQK
LYLPLPIIPATQEDGASVQGVHPSRLESGAKVRIPAEALGDGEPATVNLHWE
HGSDGAIVPRPSPGDRITLFTVRSAAANLGRLLWYITVTRAGTPTKISEKFEUR
VSTFDQRQSPYIIQVEGAENRELKLHEVPGVGCALCTLPFTFMAEGQLGLGERSAS
QGVTHILRPPTFAVYDEYSDGVIKAYLPKQLGNLDLGDKDFRLTVSASFDDGETWR
AFPIVDLKLVS"
2024..4489
/locus_tag="Pp1884"
2024..4489
/locus_tag="Pp1884"
/note="identified by Glimmer2; putative"


```

QY      121  CGAGGC 126
      |  |||||
Db      160936  CAAGGC 160931

RESULT 8
AC121979
LOCUS   AC121979          155274 bp      DNA      linear      ROD 13-NOV-2003
DEFINITION Mus musculus BAC clone RP24-289L14 from chromosome 9, complete
sequence.
ACCESSION AC121979
VERSION   AC121979.3   GI:24476151
KEYWORDS  HTG.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 155274)
           McLeellan, M., Cotton, M., Doebber, A. and Schatzkamer, K.
           The sequence of Mus musculus BAC clone RP24-289L14
JOURNAL   Unpublished (2001)
AUTHORS   Wilson, R.
TITLE     Sequencing of Mus musculus
JOURNAL   Unpublished (2001)
AUTHORS   McPherson, J.D. and Waterston, R.H.
REFERENCE 3 (bases 1 to 155274)
           Direct Submission
JOURNAL   Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
AUTHORS   McPherson, J.D. and Waterston, R.H.
REFERENCE 4 (bases 1 to 155274)
           Direct Submission
JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
AUTHORS   McPherson, J.D. and Waterston, R.H.
REFERENCE 5 (bases 1 to 155274)
           Direct Submission
JOURNAL   Submitted (02-NOV-2002) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
AUTHORS   Wilson, R.
REFERENCE 6 (bases 1 to 155274)
           Direct Submission
JOURNAL   Submitted (13-NOV-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
           On Nov 2, 2002 this sequence version replaced gi:22476161.
COMMENT   ----- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu
           Contact: submissions@watson.wustl.edu
           ----- Summary Statistics
           -----
           Center project name: M_BB0289L14
           -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see

```

http://genome.wustl.edu

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES

source

Location/Qualifiers

1..155274
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="9"
 /map="9"
 /clone="RP24-289L14"
 /clone_lib="RPCI-24"

repeat_region
 1..165
 /rpt_family="B2"

repeat_region
 178..287
 /rpt_family="Alu"

repeat_region
 288..309
 /rpt_family="B4"

repeat_region
 437..551
 /rpt_family="B2"

repeat_region
 568..732
 /rpt_family="B2"

repeat_region
 788..936
 /rpt_family="B4"

repeat_region
 1010..1081
 /rpt_family="MIR"

repeat_region
 1104..1289
 /rpt_family="MaLR"

repeat_region
 1130..1302
 /rpt_family="Alu"

repeat_region
 4855..4982
 /rpt_family="Alu"

repeat_region
 4876..4987
 /rpt_family="B4"

repeat_region
 5144..5254
 /rpt_family="Alu"

repeat_region
 5316..5565
 /rpt_family="B4"

repeat_region
 5586..5711
 /rpt_family="Alu"

repeat_region
 5712..5786
 /rpt_family="B4"

repeat_region
 5788..5975
 /rpt_family="B2"

tRNA
 5793..5865
 /product="tRNA-Cys"
 /note="Likely pseudogene (HMM Sc=39.91 / Sec struct

SC=-11.22)"

repeat_region
 7045..7194
 /rpt_family="Alu"

repeat_region
 7670..7741
 /rpt_family="Alu"

repeat_region
 8319..8498
 /rpt_family="7SLRNA"

repeat_region
 8388..8550
 /rpt_family="B4"

repeat_region
 8826..9027
 /rpt_family="B2"

tRNA
 8832..8904
 /product="tRNA-Ser"

SC=-6.89)"

repeat_region
 9947..10098
 /rpt_family="B4"

repeat_region
 10110..10490
 /rpt_family="MaLR"

```
repeat_region 11038..11277
/rpt_family="B2"
tRNA 11043..11114
/product="tRNA-Ser"
/note="likely pseudogene (HMM Sc=33.19 / Sec struct
Sc=-10.52)"
repeat_region 11238..11348
/rpt_family="B2"
repeat_region 11348..11466
/rpt_family="B2"
repeat_region 11349..11370
/rpt_family="B4"
repeat_region 11472..11624
/rpt_family="RMER6B"
/rpt_family="B2"
repeat_region 11625..11800
/rpt_family="B2"
repeat_region 11810..11946
/rpt_family="Alu"
repeat_region 11959..12092
/rpt_family="RMER6B"
repeat_region 12093..12290
/rpt_family="B2"
repeat_region 12291..12465
/rpt_family="RMER6B"
repeat_region 12916..13104
/rpt_family="B2"
repeat_region 13112..13274
/rpt_family="ERVK"
repeat_region 14294..14440
/rpt_family="Alu"
repeat_region 14507..14582
/rpt_family="Alu"
repeat_region 15774..16035
/rpt_family="RMER6B"
repeat_region 16035..16182
/rpt_family="Alu"
repeat_region 16183..16352
/rpt_family="RMER6B"
repeat_region 16353..16501
/rpt_family="Alu"
repeat_region 16502..16576
/rpt_family="RMER6B"
repeat_region 17406..17489
/rpt_family="L2"
repeat_region 17585..17934
/rpt_family="ERVL"
repeat_region 17672..17818
/rpt_family="Alu"
repeat_region 17819..18242
/rpt_family="MaLR"
repeat_region 18243..18309
/rpt_family="L2"
repeat_region 18767..18948
/rpt_family="B2"
repeat_region 18962..19079
/rpt_family="B4"
repeat_region 19415..19480
/rpt_family="B2"
tRNA 19420..19492
/product="tRNA-Ser"
/note="likely pseudogene (HMM Sc=20.93 / Sec struct
Sc=-0.89)"
```

```
Query Match 18.6%; Score 33; DB 10; Length 155274;
Best Local Similarity 51.7%; Pred.No. 48;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 3 CTTATCGTGACGAGCGCGCTTCTCGTGTGTTGCTAACCGGCGGCGGCAACTCCTTA 62
DB 45418 CTTAATGTGATCCCCCCCCCTTTCTTCGTAGCTGACGGGCGGCGGCTTT 45477
QY 63 TCGGAACGAGCGCGCTTCATACGCGCGGCTTATCTATGCGGTGACCGGACGACG 122
```

```
Db 45478 CCTACACAGCTCGCACTCTCTCTGGGGGGGCGACAGGCTCGCCCTGCCACACAGG 45537
QY 123 AGGCGCGCTCCCGCTTATCGCGCC 147
Db 45538 GACAGAGCTCCCGCTTCCCTTCC 45562

RESULT 9
CR361569/c
LOCUS CR361569.5 179571 bp DNA linear HTG 11-OCT-2004
DEFINITION Danio rerio clone CH211-99120, WORKING DRAFT SEQUENCE, 12 unordered
pieces.
ACCESSION CR361569
VERSION CR361569.5 GI:54035037
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
McLaren,S.
Direct Submission
Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Oct 11, 2004 this sequence version replaced gi:5159229.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zc99120
----- Summary Statistics
Assembly program: XCAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 175712 bases at least Q40
Consensus quality: 176324 bases at least Q30
Consensus quality: 176923 bases at least Q20
Insert size: 178471; sum-of-contigs
Insert size: 179905; 5.5% error; agarose-fp
Quality coverage: 7.91x in Q20 bases; sum-of-contigs Quality
coverage: 8.06x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 18211: contig of 18211 bp in length
* 18212 18311: gap of 100 bp
* 18312 33141: contig of 14830 bp in length
* 33142 33241: gap of 100 bp
* 33242 37691: contig of 4450 bp in length
* 37692 37791: gap of 100 bp
* 37792 44623: contig of 6832 bp in length
* 44624 44723: gap of 100 bp
* 44724 90384: contig of 45661 bp in length
* 90385 90484: gap of 100 bp
* 90485 98331: contig of 7847 bp in length
* 98332 98431: gap of 100 bp
* 98432 102355: contig of 3824 bp in length
* 102356 102356: gap of 100 bp
* 102356 111881: contig of 9526 bp in length
* 111882 111881: gap of 100 bp
* 111882 114150: contig of 2169 bp in length
* 114151 114250: gap of 100 bp
* 114251 134658: contig of 20408 bp in length
* 134659 134758: gap of 100 bp
```

* 134759 154564: contig of 19806 bp in length
* 154565 154664: gap of 100 bp
* 154665 179571: contig of 24907 bp in length.

FEATURES

source
1. 179571
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-SP120"
/clone_lib="CHORI-211"
1. 18211
/note="assembly fragment: 01057
fragment chain: 1
clone_end: SP6
vector_side: left"
18312. 33141
/note="assembly fragment: 00502
fragment chain: 1"
33242. 37691
/note="assembly fragment: 00073
fragment chain: 1"
37792. 44623
/note="assembly fragment: 00179
fragment chain: 1"
44724. 50384
/note="assembly fragment: 02183
fragment chain: 1"
90485. 98331
/note="assembly fragment: 00265
fragment chain: 1"
98432. 102255
/note="assembly fragment: 00121
fragment chain: 1"
102356. 111881
/note="assembly fragment: 00367
fragment chain: 1"
111982. 114150
/note="assembly fragment: 00054
fragment chain: 1"
114251. 134658
/note="assembly fragment: 00751
fragment chain: 1"
134759. 154564
/note="assembly fragment: 01385
fragment chain: 1"
154665. 179571
/note="assembly fragment: 01746
fragment chain: 1
clone_end: 7
vector_side: right"

ORIGIN

Query Match 18.5%; Score 32.8; DB 2; Length 179571;
Best Local Similarity 50.6%; Pred. No. 54;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
Qy 22 AGCTTCCTGTTGCTTAACCGCAGCGGACGCACTCTTATCGGAACAGACGGCGCTC 81
Db 175074 AGCCTTTTGGTGTAAATAGATGACACACATTTTCGTTCAGTCGGTCAGCAGTCTGGGC 175015
Qy 82 CATATCACCGCGCTTATCTCATGCGGTGACCGGACGAGCGCGCGCTTAT 141
Db 175014 CAGCTAACGCTCTAGGAATCTCTCGCGCTCAGGGCAACAAACCACCTGAACCAACMA 174955
Qy 142 CGCGCCTATAATACAGCCGCAACGATCTCGTAA 177
Db 174954 CCCCCAAAATCTACAGCATGAGATTTCAGGTTAA 174919

RESULT 10
AP003376 177263 bp DNA linear PLN 31-AUG-2004
LOCUS AP003376 177263 bp DNA linear
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
AUTHORS

1
Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y.,
Antonio, B. A., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
Hamada, M., Harada, C., Hijishita, S., Honda, M., Ichikawa, K.,
Idonuma, A., Iijima, M., Ikeda, M., Ikono, M., Ito, S., Ito, T., Ito, Y.,
Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S.,
Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara, T.,
Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M.,
Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,
Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Teuji, K., Waki, K.,
Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,
Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,
Yano, M., Jiang, J. and Gojobori, T.
The genome sequence and structure of rice chromosome 1
Nature 420 (6913), 312-316 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

12447438
2 (bases 1 to 177263)
Sasaki, T., Matsumoto, T. and Yamamoto, K.
Direct Submission

Submitted (07-MAR-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)

On Jul 17, 2002 this sequence version replaced gi:20975425.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI NonRedundant Protein database, nr

(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the

corresponding DDBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.

The orientation of the sequence is from M13rev to -21M13 of the BAC
clone. This sequence of OSJNBa0014K08 clone has an overlap with
P0439E07 clone (DDBJ : AP003768) at the position 1 to 45,852 of 5'
end and an overlap with P0481E12 clone (DDBJ: AP003076) at the
position 172,063 to 177,263 of 3' end. Detailed information on
overlap and assembly quality together with annotation of this entry
is available at <http://rgp.dna.affrc.go.jp/GenomesSeq.html>.

FEATURES
source

1. 177263
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="OSJNBa0014K08"
complement(join(1396..1701,1758..2043,2410..2544,
3405..3547,5098..5199,6055..6167,6605..6690,7352..7491))
/gene="OSJNBa0014K08.1"
complement(join(1396..1701,1758..2043,2410..2544,

gene
CDS

```
3405. .3547,5098. .5199,6055. .6167,6605. .6690,7352. .7491))
/gene="OSJNBa0014K08.1"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAC05577.1"
/db_xref="GI:21902028"
/translation="MAEPDWILRRRRRQVEQILELDMELQVEEVDDAGSSSSSEVDT
FLRNTHGGGRTSALAFNTSVSLPTCDGGVLFPEAILPIRVQPRSLTAVDKAV
NHVDPAGMIGVIHHIKOLDGSSNVTRQNRFLRRRWIDADDVALLSRRSDGPGVA
YKQSCQEWMTVYNATGLALGAPSKAHSLFPGTATPQRRGGEKRRHGHDIO
AAIAHLKCGFYLEKLKGNRSQRRGGAPRRLRVPTNEEDDPPPLAPALPLGR
QWCCFCGAPAGAREVPPRLTWPRRHPCGPTTSPRRHGLDDPPRHGAPQGA
RRPPFRGFRRAAIIARRPLLNRIYAEFQLSAASAPVPRSGRPLPPRRRLATG
RHHDNRGHLRPSAAAGQIQAREGRIRPGWLRIRLLTLP"
join(8072. .8086,9178. .9387,9474. .9671,12990. .13120,
14072. .14144,14960. .15376)
/gene="OSJNBa0014K08.2"
join(8072. .8086,9178. .9387,9474. .9671,12990. .13120,
14072. .14144,14960. .15376)
/gene="OSJNBa0014K08.2"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAC05578.1"
/db_xref="GI:21902029"
/translation="MRSNPLVPLPACGTGTPVPPTASLRRELALSNNLPARGG
VAGQNAATRAPAAASARVVVVSFLARRRRTRPPPPPLPSPRVFSASPP
FGLIPMRSLRRSGESGGDQAGALILSWDANYGYLDEQFCOVEDLQDEASP
NFVEVVTLPFKDSGLMSNTQALCMRSFOKVRHAYLRQKLESYFOLCSWTABE
QKILLADDLAKPSROYSMKGSATNRAIPDRSAAMSAMARERSCHARNTR
NGTRAGARPDPESRQIRVLGRSLSVDETWITSAAAFLRHGETRAPAFALGTA
GRRGEDMSVIR"
join(16191. .16218,17138. .17280)
/gene="OSJNBa0014K08.3"
join(16191. .16218,17138. .17280)
/gene="OSJNBa0014K08.3"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAC05579.1"
/db_xref="GI:21902030"
/translation="MAITVEWNGQPPSVGHCCIMRPSDVAATCRVQACSCITAAPGGT
GLLRBMAGLSSR"
join(19222. .19261,19694. .19848,19915. .20022,20628. .20832,
21203. .21300,21397. .21480,21572. .21649,21807. .21899,
21991. .22020,22277. .22421,22872. .22967,23073. .23252,
23347. .23453,23552. .23653,23956. .24011,24094. .24347,
24422. .24585,24909. .25092,25170. .25423)
/gene="OSJNBa0014K08.4"
join(19222. .19261,19694. .19848,19915. .20022,20628. .20832,
21203. .21300,21397. .21480,21572. .21649,21807. .21899,
21991. .22020,22277. .22421,22872. .22967,23073. .23252,
23347. .23453,23552. .23653,23956. .24011,24094. .24347,
24422. .24585,24909. .25092,25170. .25423)
/gene="OSJNBa0014K08.4"
/notes="contains ESTs AU065754 (S10005), AU175140 (S10005)"
/codon_start=1
/product="putative kinesin-like protein"
/protein_id="BAC05580.1"
/db_xref="GI:21902031"
/translation="MGTVNGEVEDFPAANRBAEVDLWGLLPEDLPDLSDELRD
YLINGEALCYADKLMFGLVGTGCGYASDORSNVKKFLSVVAENGLFGFGKRYEP
SMSSIVCELLAKNVATOLGHSINSTAKTPIRKKLELRETDGVLVSATPGRKYPK
SOQRPLSGKINSREHVOFGKGYTDLPAAKISEMLHSNLDNAPTQSLRNVANGILD
ESIERKGEIPIHVVHLLRNVIQETIHRGIGQADIRNQNSIIKTREDKYSKIKALE
TLVNTGENEAMNRLVAVKESKIDKRLKGQDMIRLIREKENAENIIASLHOE
MQVNRWHEQPHDLKGLVDAENYHKVLAENQKLFNEVQELKGNIRYVCRVPLPG
QKGLTALDIYIGENGELIIPASQKQEGYRFRKRVFGTHSSQAEVFSIQPLIRS
VLGDNVCIYAFQSGTSKTYMTSGPTSGREDGWVYNRALNDFDISLRKNRAYSYES
NRAVGTALNERSHSHLITVHRLDVGKNGSRGCLHLIDLAGSREVERSATG
RLKEAQHNTKLSALGDVIFSLAQNAHVYRNSKLTQVLOSSLGGQAKTLMFQINP
DIESYSETISTLKFAEYVSGVGLGARSNREGKDIKELLEQVASIKDTIARKDMEIEQ
LQLKSKSPNSMTDRNGSNLIRQSTTSGLSLPVSQOQNQLSVTLTPLADTGLAG
SVEAEADNADSDGCGVETEYSPAGASETSABRAYILILFSLFISCAKPSACADTSGE
LMIVPRLRTGT"
```

```
gene
CDS
join(27978. .28537,28795. .29152,29378. .29650)
/gene="OSJNBa0014K08.5"
join(27978. .28537,28795. .29152,29378. .29650)
/gene="OSJNBa0014K08.5"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAC05581.1"
/db_xref="GI:21902032"
/translation="MATASVTVYQRPRLPVPEFRAPPSPVTVGLTSSSSGSLAYG
ECPNSNDEITGRFLRRSARVPLRLPERAVPRKKAAWAPVIVDRLIALPEAGGPV
AEALRSAAVAGCQVUGGVDRSLVSAALBHVAAATAAATPEEVEVGGDNGE
EAWWSPGCGGMAGNALQSGASHFRNTADALFQLEQTATKINDVLRGAVATQS
TAGADTNSLGLICIRHRRRRQDRSGSGPIRHDDILRMLVRSKSRALAHLLCPGA
SAFHIFSRGWSRFRPMGDGAVVTVGDLQACNGGLYKSVAKGPSNDLDRNGGDT
GVASAEFYCFPSAGTAGKASEVLSADAGKIIPLNLFQPMVAACLVLGVHFLLSLHS
TWLL"
complement(join(31081. .31242,31391. .31579,31787. .31815,
31903. .31993,32132. .32306,32410. .32549,32638. .32770,
33508. .33574,34086. .34243,34332. .34405,34496. .34609,
35132. .35437))
/gene="OSJNBa0014K08.6"
complement(join(31081. .31242,31391. .31579,31787. .31815,
31903. .31993,32132. .32306,32410. .32549,32638. .32770,
33508. .33574,34086. .34243,34332. .34405,34496. .34609,
35132. .35437))
/gene="OSJNBa0014K08.6"
/notes="contains ESTs C72949 (E2532), C72949 (E2532)"
/codon_start=1
/product="putative casein kinase-like protein"
/protein_id="BAC05582.1"
/db_xref="GI:21902033"
/translation="MPVLSRAARRAREAQENPAAAAAPPPARRRRARRRKEPEV
AVEAAPAEEGREEIEVADLREGGEGKMEGPDGSGARSADKQAVEDEGNTTPVDT
VQVGSSPKYRVDKGLGCGQGVYGRMLANGPGASVALKFEHRTSKGNGHPPE
WQVNAAGIIGVPRVHYKQGEYVIVMTPEHYERNCFALSKILINLVYKS
TKWDACTLHVDIQRPDVFREVEYVNNLKFDPEFNACISLFDVWGNPDRIY
INTDGAQLHLVGQGRLLMEETDEQPKKIRMGPMATOWISVYNARRMPKQRIY
NVAISRVLQHIDKGNEDGLFICITSCSNLWALIMDAGTGETSOYVLESPLHREWI
MDQERNYIITALAGANNGLSVMSKGTMTQOSYKVELDFLPSGIIHQWONGY
RITATAWDOAAVLSVPRKPTDEOTELTTSAPFSQHVKKSGHLVEAISSRLLA
LVSAQAAAYVMASSCEDLAYPDYRAVVKFNGYRS"
join(39834. .39907,40580. .40722,41299. .41408)
/gene="OSJNBa0014K08.7"
join(39834. .39907,40580. .40722,41299. .41408)
/gene="OSJNBa0014K08.7"
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAC05583.1"
/db_xref="GI:21902034"
/translation="MPYVVGTTGNCSEBAIVIRSLISROKKEEYKGYKSELANPPM
AVGSLDAKSTCFKHICRDYSGERTLCVSVIENRRFAPVSLDVQNREWMVNLQCLDI
SIFPPC"
```

```
Query Match 18.3% Score 32.4; DB 8; Length 177263;
Best Local Similarity 53.1%; Pred. No. 71;
Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 9 GTGACAGACGCCAGCTTCCTGTGTGTTGCTAACCGCAGCGGACGCAACTCTTATTCGAA 68
|||||
Db 148416 GTGACAGCGGACGTACGTAATGTGTTGGAATCTCCCGCGTCAAAATTCCTTGGCGAA 148475
|||||

QY 69 CAGGACGCGCTCCATATCAGCCGCGGCTTATCTCATCGCGTACCGGACACGAGGCGC 128
|||||
Db 148476 AAGCGCTCGCTCTTGTCCGATTCCTGTCAGTTGGCGCCAGCGCTGAGGCCCC 148535
|||||

QY 129 CCGTCCCGCT 138
|||||
Db 148536 CCGGCCCTCT 148545
|||||
```

```
RESULT 11
AC134482/c
LOCUS AC134482 222256 bp DNA linear HTG 15-NOV-2002
```

DEFINITION	Rattus norvegicus clone CH230-179K4, *** SEQUENCING IN PROGRESS
ACCESSION	AC134482
VERSION	AC134482.2 GI:25007228
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 222256)
AUTHORS	Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalio,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Gant,R., Garcia,A., Garner,T., Garza,M., Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,L., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhwari,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,I.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munitasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokenleh,O., Okwuon,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Popper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quairoz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Vallas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 222256)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (27-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 222256)
AUTHORS	Rat Genome Sequencing Consortium.
TITLE	Direct Submission
JOURNAL	Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One
COMMENT	Baylor Plaza, Houston, TX 77030, USA On Nov 15, 2002 this sequence version replaced gi:23334632. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: KCOV Center clone name: CH230-179K4 ----- Summary Statistics Assembly program: Phrap; version 0.990329 Consensus quality: 190050 bases at least Q40 Consensus quality: 194119 bases at least Q30 Consensus quality: 196964 bases at least Q20 Estimated insert size: 204504; sum-of-contigs estimation Quality coverage: 6x in Q20 bases; sum-of-contigs estimation ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 2 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 107899: contig of 107899 bp in length * 107900 107999: gap of unknown length * 108000 222256: contig of 114257 bp in length. FEATURES source 1..222256 Location/Qualifiers /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clones="CH230-179K4" complement(113301..114059) note="clone boundary clone_end:Sp6 site: end_sequence:BH303478" misc_feature Query Match 18.3%; Score 32.4; DB 2; Length 222256; Best Local Similarity 53.1%; Pred. No. 71; Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0; ORIGIN QY 44 AGCGGAGCGAATCTTATCGGAACAGGAGCGCTCCATATACGCGCGGTATATCTC 103 Db 2737 AACTAGACCCCGCTCTCTCATCAGCTCAGTCACTTCTCTGGGCTGGACCCCTTATC 2678 QY 104 ATGCGCGTGACCGGACACGAGCGCCGCTTCGCCGCTTATAAATACAGCCGCG 163 Db 2677 GTTAGCCAGAGTGAGAACAGGCTCAGGTCCCTTGTACACTTCTTAAGGGAAGGTCA 2618 QY 164 AACATCTGG 173 Db 2617 AGAGTGCTGG 2608

RESULT 12

AC135655

LOCUS

DEFINITION

AC135655 270958 bp DNA linear HTG 20-NOV-2002

4 unordered pieces.

ACCESSION

AC135655.2 GI:25138854

VERSION

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

Rattus norvegicus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 270958)

REFERENCE

AUTHORS

Munzy, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Loulsegged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankweme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poinexter, A., Popovic, D., Primus, E., Pu, D.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, X., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

REFERENCE

AUTHORS

TITLE

JOURNAL

Rat Genome Sequencing Consortium.

Submitted (20-OCT-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

JOURNAL

JOURNAL

TITLE

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

JOURNAL

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:24158508.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KCM1

Center clone name: CH230-6908

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 206551 bases at least Q40

Consensus quality: 211776 bases at least Q30

Consensus quality: 215202 bases at least Q20

Estimated insert size: 208321; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 248554: contig of 248554 bp in length
 * 248555 248654: gap of unknown length
 * 248655 267829: contig of 19175 bp in length
 * 267830 267929: gap of unknown length
 * 267930 268946: contig of 1017 bp in length
 * 268947 269046: gap of unknown length
 * 269047 270958: contig of 1912 bp in length.

FEATURES

Location/Qualifiers

1..270958

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-6908"

1..1141

/note="wgs_contig"

91405..94036

/note="wgs_contig"

102157..103191

/note="wgs_contig"

242793..243887

/note="wgs_contig"

246116..248554

/note="wgs_contig"

248655..249752

/note="wgs_contig"

ORIGIN

Query Match 18.3%; Score 32.4; DB 2; Length 270958;

Best Local Similarity 53.1%; Pred. No. 70;

Matches 69; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy	44	ACCGGACGAACTCCTTATCGGAACAGGACGCGCTCCATATACGCGCGCTTATCTC	103	
Db	105760	AAC TAGACCCCGCTCCCTACGTACAGTCAATCTCTCTGGCGTGGACCCCTTCTATC	105819	
Qy	104	ATGCGCGTGACCGGACACGACGCGCGCTCCGCTTATCGGCGCTATATAATACAGCCGCGC	163	
Db	105820	GTTAGCCAGAGTGAGACAGGCTCAGTCCCTTGTACACTTCTTAAGGAAGGTCTAC	105879	
Qy	164	AACGATCTGG	173	
Db	105880	AGAGTGCTGG	105889	
RESULT 13				
AP005031		299550 bp	DNA	linear
LOCUS		Streptomyces avermitilis genomic DNA, complete genome, section		BCT 10-MAY-2003
DEFINITION		11/30.		
ACCESSION		AP005031		
VERSION		AP005031.1		GI:29606108
KEYWORDS				
SOURCE		Streptomyces avermitilis MA-4680		
ORGANISM		Streptomyces avermitilis MA-4680		
REFERENCE		Bacteria: Actinobacteria: Actinobacteridae: Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.		
AUTHORS		Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.		
TITLE		Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites		
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)		
MEDLINE		21477403		
PUBMED		11572948		
REFERENCE				
AUTHORS		Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.		
TITLE		Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis		
JOURNAL		Nat. Biotechnol. 21 (5), 526-531 (2003)		
MEDLINE		22608306		
PUBMED		12692562		
REFERENCE				
AUTHORS		Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kishida, N., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y. and Hattori, M.		
TITLE		Direct Submission		
JOURNAL		Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nite.go.jp, URL: http://www.bio.nite.go.jp/, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)		
COMMENT		This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kishida(*4), Hisaaki Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).		
		Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.		
		*1 Kitasato Institute for Life Sciences, Kitasato University		
		*2 National Institute of Infectious Diseases		
		*3 The Kitasato Institute		
		*4 National Institute of Technology and Evaluation		
		*5 School of Science, Kitasato University		
		*6 Institute of Medical Science, University of Tokyo		
		*7 RIKEN, Genomic Sciences Center		
		Following url is also available.		
		http://avermitilis.lis.kitasato-u.ac.jp.		
FEATURES		Location/Qualifiers		

1.	299550			
/organism=	"Streptomyces avermitilis MA-4680"			
/mol_type=	"genomic DNA"			
/strain=	"MA-4680"			
/db_xref=	"taxon:227882"			
/note=	"This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinius"			
complement(42..2936)				
/gene=	"nrd"			
complement(42..2936)				
/gene=	"nrd"			
/note=	"SAV2461"			
/codon_start=	1			
/transl_table=	11			
/product=	"putative vitamin B12-dependent ribonucleotide reductase"			
/protein_id=	"BAC70172.1"			
/db_xref=	"GI:29606109"			
/translation=	"MTETASGPARGSRKGTAKGLRIERTHTTPGVHPYDEVSEWAR DVMTNMRGVSFNEQRGVFPDFWVSNNVNTSKYPRGAVGTPQREVSLRLIDRI VKTRYKAGEDYKYPASPADAEIFEHELAYALLHQIFSNFSPVFNVTGTPQPVASCF ILAVDDSMESILDMYKEEGMIFKGGAGLNLGRISSELLSSGGNAGSPVSMRGA DASAGTIKSGCATRRAAKWVILDVHPDIEDFETKVKEEIKRALRDAGFMDLGGD DITSVOYONANNSVRVNDTFMKAVEGCKFGLTSRMTGEVIEVDKSLFRKMAEAAW ACADPGIQDDTJNHMHTCPESGRINGSNPSSEYMHLDNTSCNLASLMLAKFLKDDGK TGHAYSDGGRALAGAITSLMTGTSYKRSAAELAAVGPYDGYARNAQHORVMKQHS ANGVAVRDLDLTP IWAATAEWQDVLHLGKNGFRNAQSVIAPTGTIGLMSCDTT GLPDLALVKFKLVGGGSMQIVNGTVPQALRLRGYQEEQIEAIVAHIAENGVIDAP GLKHEHYEVPDCAMGERSISAMGHVRMAAIOPMISGALSKTNLNPATVEDVBEVY PEAKMGVKALAIYRDNCKVQPLSAKKKETEKAETAKTEATIREAVEKVVEVRPV KRLLKPRGPTTFTVCGAGGYMTANSYPDDGLGEVFLKMSKQSGTLAGMMDASIAV SVGLQGVPLETYVSFTNMRFPAGMTDDPDMVMAQSIVDYIFRRALDPLPPETRS ALGTHSAERORHLETGSYFSDDDMDVMDVLAQSAAPRAQELKAVATPKAEVAAVPA PKQAHTSAELVEMOLGIQADAPLCFSCCTKMQRAGSCYICEGCGSTSGCS"			
complement(3086..3649)				
/note=	"SAV2462"			
/codon_start=	1			
/transl_table=	11			
/product=	"putative regulatory protein"			
/protein_id=	"BAC70173.1"			
/db_xref=	"GI:29606110"			
/translation=	"MHCPCRHDPDSRVDSRTTDDGTSIRRRQCPDCSRFTTVEIC SLHWKRSQVTEPFSTRKVIINGVRKACQGRPVTEDALAQIQORVEEAVRATGSLRTT HDVGLAILGPRLDILVYLRFAFVSIFRAFSLEDFEAAIAELREEQRERPAVDDDEHE 4168..4947			
/gene=	"lexA"			
4168..4947				
/note=	"lexA"			
/note=	"SAV2463"			
/codon_start=	1			
/transl_table=	11			
/product=	"putative SOS regulatory protein LexA"			
/protein_id=	"BAC70174.1"			
/db_xref=	"GI:29606111"			
/translation=	"MTTADSATITADRSQGRLEPVPVHMANEATSHGPKRSLRQRP GPGRSSGLDRQRRIEIVDSVQRGYPFSMRREIQAVGLSSTSSVAHQLMALERK GFRLRDPHPRAYEVRGSDQSVQPTDTAGKPAASVPLVGRIAAGGPVLAEESEVED FVLPRQLVGDGELFVLKVVGDSMIEAAICDGMVTVRRQFPAENGDIIVAMLDEGAT VKRFRDGHVWLLPHNSAYQPIPGDEATILGKVVALRRV"			
complement(5110..7143)				
/gene=	"ding"			
complement(5110..7143)				
/gene=	"ding"			
/note=	"SAV2464"			
/codon_start=	1			
/transl_table=	11			
/product=	"putative ATP-dependent helicase"			
/protein_id=	"BAC70175.1"			
/db_xref=	"GI:29606112"			
/translation=	"MTKPSLPPELLHAAVAVTAVGGTFRPQVMTAEVAAEAIIDSSHLV			

Db		37	TAAGATCGGCC	26	
RESULT 15	AY327402/c				
LOCUS	DEFINITION				
	AY327402	131158 bp	DNA	circular VRL 08-OCT-2003	
	Choristoneura fumiferana defective nucleopolyhedrovirus complete genome.				
ACCESSION	AY327402	AF068194	U10476	U23422	U78194
VERSION	AY327402.1	GI:37499238			
KEYWORDS					
SOURCE					
ORGANISM					
	Choristoneura fumiferana defective nucleopolyhedrovirus (CFDEFNPV)				
	Choristoneura fumiferana defective nucleopolyhedrovirus				
	Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Nucleopolyhedrovirus.				
REFERENCE	1	(bases 11500 to 13512)			
AUTHORS	Barrett,J.W., Krell,P.J. and Arif,B.M.				
TITLE	Characterization, sequencing and phylogeny of the ecdysteroid UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis viruses isolated from Choristoneura fumiferana				
JOURNAL	J. Gen. Virol.	76	(Pt 10),	2447-2456	(1995)
MEDLINE	96030854				
PUBMED	7595348				
REFERENCE	2	(bases 13243 to 14241)			
AUTHORS	Barrett,J.W., Lauzon,H.A., Mercuri,P.S., Krell,P.J., Sohi,S.S. and Arif,B.M.				
TITLE	The putative LEP-1 proteins from two distinct Choristoneura fumiferana multiple nucleopolyhedroviruses share domain homology to eukaryotic primases				
JOURNAL	Virus Genes	13	(3),	229-237	(1996)
MEDLINE	97187920				
PUBMED	9035367				
REFERENCE	3	(bases 80268 to 81693)			
AUTHORS	Li,X., Lauzon,H.A., Sohi,S.S., Palli,S.R., Retnakaran,A. and Arif,B.M.				
TITLE	Molecular analysis of the p48 gene of Choristoneura fumiferana multicapsid nucleopolyhedroviruses CFMNPV and CFDEFNPV				
JOURNAL	J. Gen. Virol.	80	(Pt 7),	1833-1840	(1999)
MEDLINE	99350016				
PUBMED	10423153				
REFERENCE	4	(bases 1 to 131158)			
AUTHORS	Li,X., Barrett,J., Pang,A., Klose,R.J., Krell,P.J. and Arif,B.M.				
TITLE	Characterization of an overexpressed spindle protein during a baculovirus infection				
JOURNAL	Virology	268	(1),	56-67	(2000)
MEDLINE	20149221				
PUBMED	10683327				
REFERENCE	5	(bases 1 to 131158)			
AUTHORS	Lauzon,H.A.M., Jamieson,P.B., Krell,P.J. and Arif,B.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUN-2003)				
	Molecular Virology, Great Lakes Forestry Centre, 1219 Queen St. E., Sault Ste. Marie, Ontario P6A 2E5, Canada				
COMMENT	On or before Oct 8, 2003 this sequence version replaced gi:1117788, gi:2581770, gi:1754838, gi:4092491.				
FEATURES	Location/Qualifiers				
source	1..131158				
	/organism="Choristoneura fumiferana defective nucleopolyhedrovirus"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:74660"				
	join(129349..131158,1..20)				
	/note="ORF 149; 1629 capsid Op2/Ac9"				
	/codon_start=1				
	/product="viral capsid associated protein"				
	/protein_id="AAQ91783.1"				
	/db_xref="GI:37499384"				
	/translation="MDRQYQSVKSLNNNAHNAIDADAFQLVAGPEAHNLKRGLEG RTVLNRLADLKLAEINADYATMQVNTIESRRHFATLMRNLNLNQLQDEHFRN VLASIVARIENLSDAVNDVEITVLGSGDFEESRYATRVQYVSTDLPPPPPLTPPI PETTQAFVAPSPPAQAFVAPSPPAQAFVAPSPPAQAFVAPSPQPTFFVAPSPQPT FVAPSPQATQPTSPRQTFAAPSPVAPSPQPTFAPTTEGLTSGADEFEFAGTSV				
CDS					
	complement (1..738)				
	/note="ORF 1; ph Op3/Ac8; major occlusion body protein"				
	/codon_start=1				
	/product="polyhedrin"				
	/protein_id="AAQ91696.1"				
	/db_xref="GI:37499297"				
	/translation="MPNYSVRPTIGRTVYVYDKNYKLGSVIKNAKKKHLLEHEEDE KHLDPDHYVADDPFLGPGKNQKLTFLKEIRNVKPDPTMKLIWNMSKGFLEWTWTRF VEDSPFIVNDQEVNDVFLVNLNRPNRNCTFKPLAHLRWDCDYVPEHEVIRIVEPSY VGMNNEYIRISLAKKGGCCPIMTHSEYTNFSFESFVNRIWIENFYKPIVIYIGTDSGEEE EILIEVSLVPKVEFAPDAPLFTGPAY"				
	complement (828..1220)				
	/note="ORF 2; Op5"				
	/codon_start=1				
	/product="unknown"				
	/protein_id="AAQ91742.1"				
	/db_xref="GI:37499343"				
	/translation="MADSTNRILRLFYRWSSKTGASLEQADLECLYDLERFVGAHWN KOIETSCKKKQAEERAEASALKVELAADRMAQAKAAPYCADDGRWSTLSQQOQLED IDRNKEIVDRILYKQLQKDCLLANNKAF"				
	complement (1247..1858)				
	/note="ORF 3; lcf-2 Op6/Ac6"				
	/codon_start=1				
	/product="late expression factor 2"				
	/protein_id="AAQ91716.1"				
	/db_xref="GI:37499317"				
	/translation="MGSAMNPAAQVSSLQNSKYLIDPNDVFCVLALSPCTVFKQGL FMESGURLRALLTAPKTEPKRAVLRHSKRVCLKACADGVSNLAKALNSLRPLCM FKMAELNNAAPRGMYRRKRFETCYLGNVSCCTCKSACLIGALLHFYKMDPKCVG EVTHLLIKABDVYKPSNCARMTKNLCPRAGMCKGNKPNCP"				
	complement (1861..2124)				
	/note="ORF 4; Op7/Ac5"				
	/codon_start=1				
	/product="unknown"				
	/protein_id="AAQ91760.1"				
	/db_xref="GI:37499361"				
	/translation="MNRPMWRTANAPVIVSNHDYDQIRRDLSLRSSVHELCTRST TGEDCNRIELSSSSNDVITKSPVIIKNSAATVQSALVCDKV"				
	2171..2688				
	/note="ORF 5; Op8/Ac4"				
	/codon_start=1				
	/product="unknown"				
	/protein_id="AAQ91737.1"				
	/db_xref="GI:37499338"				
	/translation="MASARFIKFSLRVTQGFKENIVADITHLARLALIDGKVTHTDV HRFGLDRNALVEACMAVNVQVYVDPDGTIRLNPKTIYFRVCQKCHADVAPADPHSI ARYLCAACGMVLVIDHPDVGFDTEEGVNELIEVORINAGDL"				
	2647..3657				
	/note="ORF 6; bro Ac2"				
	/codon_start=1				
	/product="baculovirus repeated ORF"				
	/protein_id="AAQ91674.1"				
	/db_xref="GI:37499275"				
	/translation="MAQIKIQGFQFGEDEVFTLRVLDRIIVKFAKDVASSLKYNICD KAIRTHVDKDYKTSFEQTIOLGSGSTNLVKRGDPLYLQPHPTLITKSGVQLIMKSK LPYAVELQEWLLEEVIPQVLCTGKYNPAIKQQQENKOLINKLVKTFSDHNTLQTL LOKTQELVYKQEFIERIVATKQIEAKDLQVTVMTDLARMYNGFOETWCKKDEIMO QKDAQVTDLVAKVIDLSGRAVQYPERKGPVLCVTWDGDTFTTALAQKAYVHNQKLK RNLSAADIVAEETTRPNTVDNWNATHRLAAKSKRTISFDSSEQDAQQFATRIKQLLNN FD"				
	complement (3686..4168)				
	/note="ORF 7; ptp-2 Op9"				
	/codon_start=1				
	/product="protein tyrosine phosphatase 2"				
	/protein_id="AAQ91731.1"				
	/db_xref="GI:37499332"				
	/translation="MYDANQIDENLFGVGGYGNNEAMLQFIKKHDIIESVILSDSDVG				

PIROALGLPAGVHHIVHYCEDEBPTCMALINAMDALVDYIYIERRINECKKILIHCHAGVSR
SATLAVYYVMKKQVSYEKALRFVNNKRNVALSDHFVFLSSRCTYRFVNNKLIHVG

CDS

complement (4155, .4676)
/note="ORF 8; ptp-1 Op10/Ac1"
/codon_start=1
/product="protein tyrosine phosphatase 1"
/protein_id="AAQ91726.1"
/db_xref="GI:37499327"
/translation="MFGKRWNYTTCGKVIETNLCFKPLKQEVFFVYVNDDEDRWT
VSNLNRORALGAVIDLNTLRYLDGAKVDAGVLYKKIRVPGQEVNEDIVQEPFDS
VQEFSGCPGMLIGVCHTGLNRTGYLCVRIIDKLHVSPTDAIARFAARGHKIERR
NYLEDLLRRHVR"

CDS

4935, .5921
/note="ORF 9; Op11/Ac11"
/codon_start=1
/product="unknown"
/protein_id="AAQ91675.1"
/db_xref="GI:37499276"
/translation="MSLCSQLLVAYYGYEYNNPHERYSESYHLRYRIVNEYILINSYANE
TSCVRDIETARRLNRGELSDDARRQLNLBEIAKRLATWYHTGEMKSCSEIOSVLA
EINCHAPLETGASRVAFFALDSLEFPNMTDSQVILMGFMHFGVGYALTHVANVF
DPTIKLDGWNKFCVLTYMYRIIRGTPAELITPLQNVYIKYIKPEYDESNALAMG
DYGFCGIGKDHFAQHAKRSYIIPFQMRGEVYADRFPCFSVKKDFGRQCKETYK
DIQLQVDILHSHAMTDKQNALFDLLCCNNASDIDVDYDYIVKKFYNIAY"
6450, .6723

repeat_region

/note="4 copies of a 30 bp imperfect palindromic sequence;
location similar to Ac hrla"
/function="enhancer; replication origin"
complement (6723, .7871)
/note="ORF 10; Op20/Ac20"
/codon_start=1
/product="unknown"
/protein_id="AAQ91663.1"
/db_xref="GI:37499264"

CDS

/translation="MYRILIVFLFALLIYVWPFYQAYQHIOAQHDYNDLTDRMD
YESVMRRHHVPMELPSQVDFDNLGLAGETLKMSPVLYTEIDILFFDCTQVCE
DPGASYFFVGEGDYVNGHKLAVGGYCTNSVPRDCNRETSSVILMSLNQWTCIABDP
RYFAGTSNMTQLAGHFDRILPGQSDRNVLFDRLGREVNAVNTFRSSWDELLEDG
TRRFPMRCNARNNNLMFNPLNPLECLPNVCTNNVYVHTSVRPNFETGCDGDEA
VTRVHVVGDRSSVACISIVGLDPTTASHRFVECVNTYTSIGNFSNNKLLCPSDTF
DSNTDAAPAEVPGSPGLSGNGLDEPFRFLDTRSRIRYNDVRGLIN"
7898, .8863
/note="ORF 11; arif Op19/Ac21"
/codon_start=1
/product="actin rearrangement inducing factor"
/protein_id="AAQ91678.1"
/db_xref="GI:37499279"

CDS

/translation="MLAQINYLQVLHAALYITILIAFVFLMGTINIKYAFLELN
DNHGSVILNSVLTAFLGCPVFTTTTWAIRYKFLCYKRAEMHSNFYMKTIISLAHVMA
LVCWTLFVVFQPIYKNGHVPVLDARYDRYDRNSLCWSNI VSDTYEVHDTNAIRTDEN
CVYRHDFVKCVKCRMEVRHDEPTVENONOCALIMVMNTAVLOPFNNYVORKEMRYK
PTPVKTLXPESAPLEQDTADEEEQQSFRMLEIISSEPRVQFPPESSSLDRSSPP
FIVQSSSPNGSDGIDYDIFQPFYSVENKVVCKYLCKTHATLCA"
complement (9013, .9327)
/note="ORF 12; Op18/Ac19"
/codon_start=1
/product="unknown"
/protein_id="AAQ91753.1"
/db_xref="GI:37499354"

CDS

Query Match 18.1%; Score 32; DB 14; Length 131158;
Best Local Similarity 87.5%; Pred. No. 94;
Matches 35; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 137 CTTATCGCGCCTATAATACAGCCCGCAACGATCTGGTAA 176

DB 125416 CTAATAACGCTATAATAACAGCCCGCTGGTAA 125377

Search completed: October 24, 2005, 20:18:08
Job time : 1081.29 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:41:23 ; Search time 1277.71 Seconds
(without alignments)
5273.001 Million cell updates/sec

Title: US-09-896-888A-1_COPY_351_527

Perfect score: 177

Sequence: 1 gctctatcgtacaggagcgc.....gcccgcaacgatctggtaaa 177

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: gb_est1:*
- 2: gb_est2:*
- 3: gb_hlc:*
- 4: gb_est3:*
- 5: gb_est4:*
- 6: gb_est5:*
- 7: gb_est6:*
- 8: gb_gss1:*
- 9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	33.8	19.1	507	8	AQ297217	HS_2266_A
2	33.6	19.0	343	5	BY342594	BY342594
3	33.6	19.0	402	5	BY077721	BY077721
4	33.2	18.8	931	7	CV068875	f2_new.ch
5	33	18.6	348	5	BY338426	BY338426
6	33	18.6	353	5	BY337247	BY337247
7	33	18.6	354	5	BY013584	BY013584
8	33	18.6	355	5	BY340467	BY340467
9	33	18.6	360	5	BY343718	BY343718
10	33	18.6	362	5	BY339252	BY339252
11	33	18.6	371	5	BY083658	BY083658
12	33	18.6	611	7	CF916169	CF916169
13	33	18.6	621	5	BQ620008	TaLr1138G
14	33	18.6	624	7	CF913548	CF913548
15	33	18.6	635	2	AW319335	un11f04.Y
16	33	18.6	641	1	AA989838	uc79f07.Y
17	33	18.6	657	7	CF915946	B0987F04
18	33	18.6	669	7	CF174622	B0941F09
19	33	18.6	973	6	CA490927	AGENCOURT
20	32.8	18.5	638	9	CE066545	ligr-gss-
C 21	32.6	18.4	786	9	CL421872	2MMBB043
22	32	18.1	327	6	BY789720	BY789720
23	32	18.1	1297	5	BQ645920	AGENCOURT
C 24	31.8	18.0	553	7	CR540072	DKF2P459D

C 25	31.8	18.0	1076	5	BX463482	BX463482
26	31.6	17.9	1011	9	CL494562	CL494562
C 27	31.6	17.9	1310	5	BQ720271	AGENCOURT
28	31.4	17.7	344	5	BY340052	BY340052
29	31.4	17.7	374	2	BB843476	BB843476
C 30	31.4	17.7	681	9	CG295117	CG295117
31	31.2	17.6	693	9	CL190317	CL190317
32	31.2	17.6	2199	3	CNS0A7F6	Arabidops
33	31	17.5	829	7	CK192985	CK192985
34	30.8	17.4	394	7	CO068301	MDfrt3029
C 35	30.8	17.4	446	7	CO320619	CO320619
36	30.8	17.4	453	7	CO867977	CO867977
C 37	30.8	17.4	639	7	CF886875	CF886875
38	30.8	17.4	653	7	CF891128	CF891128
39	30.8	17.4	660	6	CB853868	CB853868
C 40	30.8	17.4	676	2	BE275414	BE275414
41	30.8	17.4	805	8	CC437710	PUHJS21TD
C 42	30.8	17.4	842	5	BUI40827	BUI40827
43	30.6	17.3	1074	8	CC245979	CC245979
44	30.4	17.2	352	5	BY342724	BY342724
C 45	30.4	17.2	647	9	AG090615	Pan trogl

ALIGNMENTS

RESULT 1
AQ297217
LOCUS AQ297217 507 bp DNA linear GSS 15-DEC-1998
DEFINITION HS_2266_A1_A08_MR_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2266 Col=15 Row=A, genomic survey sequence.
ACCESSION AQ297217 GI:4014396
VERSION AQ297217
KEYWORDS GSS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 507)
AUTHORS Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Mahairas,G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Koller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
JOURNAL Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
PUBMED 99380589
COMMENT 10449784
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2266 row: A column: 15
Class: BAC ends
High quality sequence stop: 507.
Location/Qualifiers
1..507
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2266 Col=15 Row=A"
/sex="male"

FEATURES
source
1..507
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="plate=2266 Col=15 Row=A"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 19.1%; Score 33.8; DB 8; Length 507;
Best Local Similarity 53.4%; Pred. No. 5.1;

```

Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
Qy 5 TATCGTGACAGCGCCAGCTCTCTGTTCTTAACCCGACGCGGACCACTCCTTATC 64
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 TTTCTTTAGTCACCCCTACCACTCTGTTGGCATGAGAATCCGCTATCAACCCCTTATG 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 65 GGAACAGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGCGTGACCGGACGAG 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 GGCACAGCCCTTCAACACGCTCGTGCAGAGAAACACATGACGGCATTTCAACCAAT 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 125 GCGCCGCGCCCGC 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 GTCACGTGCTCTC 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 2
 BY342594 343 bp mRNA linear EST 12-DEC-2002
 LOCUS BY342594 RIKEN full-length enriched, whole joints Mus musculus cDNA
 DEFINITION clone L230041A03 5', mRNA sequence.

ACCESSION BY342594
 VERSION BY342594.1 GI:26572082
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 343)

REFERENCE
 AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, K.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
 Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,
 Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
 Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
 Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
 Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
 Verardo, R., Wagner, L., Walestedt, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
 Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
 Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
 Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)
 MEDLINE 22354683
 PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Tissues were provided by Vassilis Aidinis (Biomedical Sciences
 Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
 Fleming street 16672 Vari, Greece) whose assistance we gratefully
 acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.

FEATURES
 source Location/Qualifiers
 1..343
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone_lib="RIKEN full-length enriched, whole joints"
 /tissue_type="whole joints"
 /clone_lib="RIKEN full-length enriched, whole joints"

ORIGIN

Query Match 19.0%; Score 33.6; DB 5; Length 343;
 Best Local Similarity 51.7%; Pred. No. 5.9;
 Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
 Qy 3 CTTATCTGTCAGACGCGCCAGCTTCTGTGTCTTAACCGGCGGACGCACTCCTTA 62
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 99 CTTAATGTGATCCCGCCCGCTTCTCTTTCTCCGTAGCTGACGCGGAGACGGCTTT 158
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Qy 63 TCGGAACAGGACGCGCTCCATATCAGCGCGCTTATCTATCGCGGTGACCGACGACG 122
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 159 CCTCACCAGCTCGCACTCTCTCGGGGCGGACGCTCGCGTCCCGTCCCGCCACCG 218
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Qy 123 AGGCGCGCTCCCGCTTATCGCGCC 147
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
 Db 219 GACAGAGCGTCCCGCTTCCCTCTCC 243
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3

BY077721 402 bp mRNA linear EST 06-DEC-2002
 LOCUS BY077721 RIKEN full-length enriched, adult male liver Mus musculus
 DEFINITION cDNA clone K63011C08 5', mRNA sequence.

ACCESSION BY077721
 VERSION BY077721.1 GI:26179135
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 402)

REFERENCE
 AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, K.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C.,
 Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H.,
 Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
 Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
 Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,

Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konaaya, A., Kurochkin, I.V., Lee, Y., Lennard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nakata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shinada, K., Sulcane, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
Nature 420, 563-573 (2002)

MEDLINE
22354683

PUBMED
12466851

COMMENT
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES
Location/Qualifiers
1..402
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K630011C08"
/sex="male"
/tissue_type="liver"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male liver"

ORIGIN
Query Match 19.0%; Score 33.6; DB 5; Length 402;
Best Local Similarity 51.7%; Pred. No. 5.9;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

3 CTTATCGTACAGGCGCGCTTCTCTGTTGCTTAACGGAGCGGACGCACTCTTCTTA 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

89 CTTAATGTGATCCCCCCCCCTTCTTCTCCGTAGCTGAGCGGCGAGACGCCCTTT 148
63 TCGGAACAGGAGCGCCTCCATATCAGCGCGCGTTCATCTCATGCGCGTGACCGGACGC 122
149 CCTCACCAGTGCCTCTCTCGGGGGCGACGGCTCGCCCTGCCCTGCCACCAACG 208
123 AGGCGCGCTCCCGCTTATCGGCC 147
209 GACAGAGCGTCCCGCTTCCCTCTCC 233

RESULT 4
CV068875 931 bp mRNA linear EST 24-AUG-2004
f2.new_chopped.fasta.Contig488 Preamplified custom cDNA library in
pCMVSPORT6.1 (ResGen, Invitrogen Inc.) Emiliania huxleyi cDNA, mRNA
sequence.
CV068875
CV068875.1 GI:51532039
EST.
Emiliania huxleyi
Emiliania huxleyi
Eukaryota: Haptophyceae; Isochrysidales; Emiliania.
REFERENCE 1 (bases 1 to 931)
AUTHORS Wahlund, T.W., Zhang, X. and Read, B.A.
TITLE Expressed Sequence Tag Profiles from Calcifying and Non-Calcifying
Cultures of Emiliania huxleyi
JOURNAL Micropaleontology (2004) In press
COMMENT Contact: Betsy Read
Department of Biological Sciences
California State University San Marcos
333 S. Twin Oaks Valley Road, San Marcos, CA 92096-0001, USA
Tel: 760 750 4129
Email: bread@csusm.edu.

FEATURES
Location/Qualifiers
1..931
/organism="Emiliania huxleyi"
/mol_type="mRNA"
/strain="1516"
/db_xref="taxon:2903"
/dev_stage="Late log growth phase"
/clone_lib="Preamplified custom cDNA library in
pCMVSPORT6.1 (ResGen, Invitrogen Inc.)"
/note="Emiliania huxleyi grown in Artificial Seawater
(Guillard's F/2 media)."

ORIGIN
Query Match 18.8%; Score 33.2; DB 7; Length 931;
Best Local Similarity 56.4%; Pred. No. 8.2;
Matches 62; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

39 ACCGCGAGCGGACGCACTCTTATCGGAACAGGAGCGCCTCCATATCAGCGCGCGTT 98
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
425 ACCTGCTCGGCGCGCGCGGTCTCGCAACAGGCGAGTCCGCGCAACAGTGCAGG 484
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

99 ATCTCATCGGTGACCGGACAGGAGCGCGCTTATCGCGCTTATCGCGCT 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 GGCGCGTGCAGGAGCGCACCTCGGGGTGGGTGTCGCCCGAATCGGCTT 534
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
BY338426 348 bp mRNA linear EST 11-DEC-2002
LOCUS BY338426 RIKEN full-length enriched, whole joints Mus musculus cDNA
clone L230014G21 5', mRNA sequence.
BY338426
BY338426.1 GI:26534820
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota: Chordata; Craniata: Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 348)

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaïdo, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gough, J., Grimmond, S., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, J., H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nunata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Itoh, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suihō-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, W., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

ORIGIN

Query Match 18.6%; Score 33; DB 5; Length 348;
Best Local Similarity 51.7%; Pred. No. 9.2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3 CTTATCGTGACAGACGCCAGCTTCCTGTGTGTGTTAAACCGCGGAGCAATCTCTTA 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db CTTAATGTGATCTCCGCCCTTTTCTCTTCCTGCTAGCTGACGCGGAGCGCGCTTT 158

QY 63 TCGAAGACAGGAGCGGCTTCATATACGCGCGGCTTATCTCATGCGCGTACCGGACAGC 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db CCTCACCAGCTCGCACATCTCTTCGCGGCGCACGGCTCGCGCTGCCCTGCCACCGG 218

QY 123 AGGCGCGCTCCCGCTTATCCCGCG 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db GACAGAGCGTCCCGCTTCCCTCTCC 243

RESULT 6

BY337247

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY337247 353 bp mRNA linear EST 11-DEC-2002
BY337247 RIKEN full-length enriched, whole joints Mus musculus CDNA
clone L230006D15 5', mRNA sequence.

BY337247.1 GI:26532790

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 353)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaïdo, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gough, J., Grimmond, S., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawai, J., H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Nunata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Itoh, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suihō-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

FEATURES

source

Location/Qualifiers
1..348
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"

Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Wataniki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Tissues were provided by Vassilis Aidinis (Biomedical Sciences Research Center 'Al. Fleming' Institute of Immunology 14-16 Al. Fleming street 16672 Vari, Greece) whose assistance we gratefully acknowledge.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
 1. 353
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="L2300006D15"
 /tissue_type="whole joints"
 /clone_lib="RIKEN full-length enriched, whole joints"
 Query Match 18.6%; Score 33; DB 5; Length 353;
 Best Local Similarity 51.7%; Pred. No. 9.2;
 Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
 QY 3 CTTATGTCACAGGACCGCTCTCTGTGTCTAAGCGGACGCGAGCAACTCCCTTA 62
 |||||
 Db 99 CTTATGTCATCCCCCCCCCTTCTCTTCGTAGCTGACGGGACGACGCGCTTT 158
 |||||
 QY 63 TCGGAACAGGACGCGCTCCATATACGCGCGGTTATCTATGCGGTGACCGGACACG 122
 |||||
 Db 159 CCTCACCAGCTCGCACTCTCTCGGGGGCGACGGCTCGCGCTGCCCTGCCACACCG 218
 |||||
 QY 123 AGGCGCCGTCGCGTTATCGCGCC 147
 |||||
 Db 219 GACAGAGCGTCCGCTTCCCTCTCC 243
 |||||

RESULT 7

BY013584
 LOCUS BY013584 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus
 DEFINITION musculus cDNA clone G730040E06 5', mRNA sequence.
 BY013584
 ACCESSION BY013584.1 GI:26073833
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 354)
 REFERENCE Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Oshino, N., Saito, R., Suzuki, H., Yamana, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,

Schönbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bratt, D., Bruscia, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Giesi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltas, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Santolin, A., Schneider, C., Semple, C. A., Setou, M., Shinada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Vexardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyshaw-Boriss, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavoian, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Harai, A., Hashizume, M., Imotani, K., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Wataniki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
 1. 354
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="G730040E06"
 /tissue_type="lung"
 /cell_line="RCB-0558 LLC"
 /clone_lib="RIKEN full-length enriched, lung RCB-0558 LLC cDNA"

ORIGIN

```

Query Match      18.6%; Score 33; DB 5; Length 354;
Best Local Similarity 51.7%; Pred. No. 9.2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 3 CTTATCGTGACAGGACGCCCTTCCTGTTGTTGCTAACCGGCGGACGCAACTCCTTA 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116 CTTAATGTGATCCCCCCCCCTTCTCTTCCTGCTAGCCTGACGCGGACGCGGCTTT 175

Qy 63 TCGGAACAGGACGGCGCTTCATATCAGCGCGCGTTCATCTATCGGCTGACCGGACAG 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 CCTCACCAGCTCGCACTTCCTCGGGGCGCACGGCTCGCGCTGCCCTGCCACCGG 235

Qy 123 AGGGCGCGCTCCCGTTCATCGCGC 147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 GACAGAGCGTCCCGTTCCTCTCC 260

RESULT 8
LOCUS BY340467
DEFINITION BY340467 RIKEN full-length enriched, whole joints Mus musculus cDNA
          clone L230028H13 5', mRNA sequence.
ACCESSION BY340467.1 GI:26569955
VERSION BY340467.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 355)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yananaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusci,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,F., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suhiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,N., Nakamura,M., Nomura,K., Numazaki,R.,

```

```

Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Vassilis Aidinis ( Bionomedical Sciences
Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
Fleming street 16672 Vari,Greece ) whose assistance we gratefully
acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES             Location/Qualifiers
     source            1..355
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="L230028H13"
                     /tissue_type="whole joints"
                     /clone_lib="RIKEN full-length enriched, whole joints"
ORIGIN
Query Match      18.6%; Score 33; DB 5; Length 355;
Best Local Similarity 51.7%; Pred. No. 9.2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 3 CTTATCGTGACAGGACGCCCTTCCTGTTGTTGCTAACCGGCGGACGCAACTCCTTA 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 CTTAATGTGATCCCCCCCCCTTCTCTTCCTGCTAGCCTGACGCGGACGCGGCTTT 158

Qy 63 TCGGAACAGGACGGCGCTTCATATCAGCGCGCTTCATCTCATCGCTGACCGGACAGC 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 CCTCACCAGCTCGCACTTCCTCGGGGCGCACGGCTCGCGCTGCCCTGCCACCGG 218

Qy 123 AGGGCGCGCTCCCGTTCATCGCGC 147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 GACAGAGCGTCCCGTTCCTCTCC 243

RESULT 9
LOCUS BY343718
DEFINITION BY343718 RIKEN full-length enriched, whole joints Mus musculus cDNA
          clone L230047A12 5', mRNA sequence.
ACCESSION BY343718
VERSION BY343718.1 GI:26573206
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 360)
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yananaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusci,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,

```

Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustinich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reid,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavalon,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Vassilis Aidinis (Biomedical Sciences
Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
Fleming street 16672 Vari,Greece) whose assistance we gratefully
acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
FEATURES
source
1. .360
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="L230047A12"
/tissue_type="whole joints"
/clone_lib="RIKEN full-length enriched, whole joints"
ORIGIN
Query Match 18.6%; Score 33; DB 5; Length 360;
Best Local Similarity 51.7%; Pred. No. 9.2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3 CTTATCGTGCAGGAGCGCAGCTTCCTGTGTGTTAAGCGAGCGGACGCAACTCCTTA 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db CTTAATGATGATCCCCCCCCCTTCCTTCTCCGTAGCTGACGCGGAGCAGCGCTTT 158
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 63 TCGGNACAGGAGCGCCTCCATATCAGCGCGCGGTATTCATCGCGGTGACCGACACG 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db COTCACCAGCTCGCACTCTCTCGGGGCGCAGCGCTCGCGCTGCGCCCTGCCACCAAG 218
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 123 AGGCCGCCGTCCCGCTTATCGCGCC 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db GACAGAGCGTCCCGTTCCTCTCC 243
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 10
BY339252
LOCUS
BY339252 RIKEN full-length enriched, whole joints Mus musculus CDNA
clone L230021B07 5', mRNA sequence.
DEFINITION
BY339252
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 362)
AUTHORS
Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Oeato,N., Saito,R., Suzuki,H., Yamanaka,I.,
Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
Batalov,S., Beisel,K.W., Blake,J.A., Bratt,D., Brusic,V.,
Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
Flotcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
Gustinich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
Ravasi,T., Reed,J.C., Reid,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
Yang,L., Yuan,Z., Zavalon,M., Zhu,Y., Zimmer,A., Carninci,P.,
Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
Rogers,J., Birney,E. and Hayashizaki,Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
22354683
12466851
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
Hayashizaki,Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Tissues were provided by Vassilis Aidinis (Biomedical Sciences Research Center 'Al. Fleming' Institute of Immunology 14-16 Al. Fleming street 16672 Vari,Greece) whose assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers
1. 362

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="L230021B07"
/tissue_type="whole joints"
/clone_lib="RIKEN full-length enriched, whole joints"

ORIGIN

Query Match 18.6%; Score 33; DB 5; Length 362;
Best Local Similarity 51.7%; Pred. No. 9.2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Oy 3 CTTATCGTGACGAGCGCAGCTTCCTGTGTGCTAACCGCAGCGGACGCACTCCTTA 62
|||||
Db 99 CTTTATGTGATCCCCCCCCCTTCCTTCCTCGTAGCTGACGGCAGGCGGCTTT 158
|||||

Oy 63 TCGGAACAGGACGGCGCTCCATATACGCGCGCGTTATCTCATCGGTCACCGGACAG 122
Db 159 CCTCACCAGCTCGCACTCTCTCGGGGCGCACGCGCTCGCGCTGCCCTGCCACCGG 218
|||||

Oy 123 AGGCGCGCTCCGCTTATCGCGC 147
Db 219 GACAGAGCGTCCCGCTTCCCTCTCC 243
|||||

RESULT 11

BY083658
LOCUS BY083658 RIKEN full-length enriched, 10 days neonate heart Mus
DEFINITION musculus cDNA clone K63004L13 5', mRNA sequence.
ACCESSION BY083658
VERSION BY083658.1 GI:26203394
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 371)

REFERENCE

AUTHORS Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A.,

Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tonita, M., Varadar, R., Wagner, L., Wahlested, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Tasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

MEDLINE

22354683

PUBMED

12466851

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

1. 371
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K63004L13"
/tissue_type="heart"
/dev_stage="10 days neonate"
/clone_lib="RIKEN full-length enriched, 10 days neonate heart"

ORIGIN

Query Match 18.6%; Score 33; DB 5; Length 371;
Best Local Similarity 51.7%; Pred. No. 9.2;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Oy 3 CTTATCGTGACGAGCGCAGCTTCCTGTGTGCTAACCGCAGCGGACGCACTCCTTA 62
|||||
Db 111 CTTTATGTGATCCCCCCCCCTTCCTTCCTCGTAGCTGACGGCAGCGCGCTTT 170
|||||

```

QY 63 TCGGAACAGGACGGCCTCCATATCAGCGCGGTATCTATGCGCGTACCGGACACG 122
Db 171 CCTCACCAGCTCGCACTCTCTCGGGGGCGACGGCTCGCGCTGCCCTGCCACACGG 230
QY 123 AGGCGCCGTCGCGCTTATCGCGCC 147
Db 231 GACAGAGCGTCCCGCTTCCCTCTCC 255

RESULT 12
CF916169
LOCUS B0991A12-5 NIA Mouse Unfertilized Egg cDNA Library (Long 1) Mus
DEFINITION musculus cDNA clone NIA:B0991A12 IMAGE:30480971 5', mRNA sequence.
ACCESSION CF916169
VERSION CF916169.1 GI:38187371
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 611)
Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
21429098
11544199
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: B0991 row: A column: 12
Seq primer: M13 Reverse
High quality sequence stop: 611
POLYA=No.

FEATURES
Location/Qualifiers
1..611
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0991A12-5"
/db_xref="taxon:10090"
/clone="NIA:B0991A12 IMAGE:30480971"
/dev_stage="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
1)"
/Note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
Site 2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
RNAs were extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer (Invitrogen):
5'-pGACAGTCTAGATCGGAGCGCGCGCCCTTTTCTTTT-3',
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker L1-SalI, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 2.5 kb. The library was

```

constructed by Yulan Piao."

```

ORIGIN
Query Match 18.6%; Score 33; DB 7; Length 611;
Best Local Similarity 51.7%; Pred. No. 9.3;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3 CTTATCGTGACAGGACGCCCTTCTCTGTGTGTACCGGCGGACGCAACTCTTTA 62
Db 49 CTTAATGTGATCCCGCCCTTCTCTCGTAGCTGACGGGCGAGCGCGCTTT 108

QY 63 TCGGNACAGGACGGCCTCCATATCAGCGCGGTATCTATGCGCGTACCGGACACG 122
Db 109 CTTACACAGCTCGCACTCTCTCGGGGGCGACGGCTCGCGCTGCCCTGCCACCGG 168

QY 123 AGGCGCCGTCGCGCTTATCGCGCC 147
Db 169 GACAGAGCGTCCCGCTTCCCTCTCC 193

RESULT 13
B0620008
LOCUS B0620008
DEFINITION B0620008 621 bp mRNA linear EST 28-JUN-2002
TAb1138G03F TAb11 Triticum aestivum cDNA clone TAb1138G03F, mRNA
sequence.
ACCESSION B0620008
VERSION B0620008.1 GI:21625087
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
1 (bases 1 to 621)
Cloutier, S., Dong, G. and Walsh, A.
Wheat functional genomics - Thatcher Lr1 cDNA library
Unpublished (2001)
Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.gc.ca
was cloned directionally, not all sequences generated with reverse
primer where from the 5' end (same with forward primer and 3' end).
Average insert size is >2.2kb
Plate: 138 row: G column: 03
Seq primer: M13 Forward.
Location/Qualifiers
1..621
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Thatcher Lr1"
/db_xref="taxon:4565"
/clone="TAb1138G03F"
/tissue_type="leaf tissue"
/dev_stage="14 Days old"
/lab_host="E. coli XL0LR"
/clone_lib="TAb11"
/Note="Vector: Lambda ZapII; mass excised in plasmid
vector pBK-CMV (Stratagene).; Site 1: EcoRI; Site 2: XhoI;
mRNA obtained from wheat N11 Thatcher Lr1 24 hours after
inoculation with leaf rust pathogen Puccinia tritici
race BBB carrying the avirulence gene Avr1."

ORIGIN
Query Match 18.6%; Score 33; DB 5; Length 621;
Best Local Similarity 51.7%; Pred. No. 9.3;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 21 CAGTTCTCTGTGTCTNACCGGCGGCGGACGCACTCTTATCGGACACGACGCGCT 80
Db 435 CACCTTCGGGTGGCGCTTCGCGCGCGCCCTTCCGAGAGCCCTTGCTCCCTTGC CGCGCC 494

```

```

Qy 81 CCATATGAGCGCGGTTATCTCATGCGGTGACCGACAGCGCGCCCTCCGGCTTA 140
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 CGACATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
Qy 141 TCGCGCCCTATAAATACAGCCGCCAA 165
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
555 CGGCACGAGCGCGGCACAGAGCAA 579

RESULT 14
CF913548 624 bp mRNA linear EST 05-NOV-2003
LOCUS B0950D02-5 NIA Mouse Unfertilized Egg cDNA Library (Long 1) Mus
DEFINITION Musculus cDNA clone NIA:B0950D02 IMAGE:30477061 5', mRNA sequence.
ACCESSION CF913548
VERSION B0950D02
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
REFERENCE Construction of long-transcript enriched cDNA libraries from
AUTHORS submicrogram amounts of total RNAs by a universal PCR amplification
TITLE method
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
MEDLINE 21429098
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6920, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: B0950 row: D column: 02
Seq primer: M13 Reverse
High quality sequence stop: 624
POLYA-No.

FEATURES
source
1. .624
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0950D02-5"
/db_xref="taxon:10090"
/clone="NIA:B0950D02 IMAGE:30477061"
/dev_stage="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
1)"
/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
Site_2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
RNAs were extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen:
5'-TGACTAGTCTAGATCGGAGCGCGCCCTTTTCTTTT-3'],
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker Lu-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into the SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 2.5 kb. The library was

```

constructed by Yulan Piao."

```

ORIGIN
Query Match 18.6%; Score 33; DB 7; Length 624;
Best Local Similarity 51.7%; Pred. No. 9.3;
Matches 75; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 3 CTATATCGTACAGAGCGCGCTCCATATACAGCCGCGCTTATCTCATGCGGTGACCGACAGCGCGCTTCTTA 62
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
33 CTTAATGTGATCCCGCCCTTTCCTTTTCCTGTAGCCTTGACGCGCGACAGCGCGCTTT 92

Qy 63 TCGGAACAGAGCGCGCTCCATATACAGCCGCGCTTATCTCATGCGGTGACCGACAGCGCGACG 122
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
93 CCTCACAGAGCTCGCACTCTCTCTCGGGGCGGACAGGCTCGCGGTGCGCCCTGCCACACG 152

Qy 123 AGGCGCCCGTCCCGCTTATCGCGCC 147
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 GACAGAGCGTCCCGCTTCCCTCTCC 177

RESULT 15
AW319335 635 bp mRNA linear EST 25-JAN-2000
LOCUS unllf04.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
DEFINITION IMAGE:2395039 5' similar to TR:Q9YGP5 Q9YGP5 RRM-TYPE RNA-BINDING
PROTEIN HERMES. ; mRNA sequence.
ACCESSION AW319335
VERSION AW319335.1 GI:6748879
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
REFERENCE The WashU-NCI Mouse EST Project 1999
AUTHORS Unpublished (1999)
TITLE Contact: Marra M/WashU-NCI Mouse EST Project 1999
JOURNAL Washington University School of Medicine
COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:1013187
Seq primer: custom primer used
High quality sequence stop: 305.

FEATURES
source
1. .635
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:2395039"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Sugano mouse kidney mkia"
/note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [TGTGGCCTACTG], digested
and cloned into distinct draIII sites of the pME18S-FL3
vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was
performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo)

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:39:18 ; Search time 193.335 Seconds
(without alignments)
5419.578 Million cells updates/sec

Title: US-09-896-888A-1_COPY_351_527

Perfect score: 177

Sequence: 1 gcttatcgtgacagcgc.....gcccgcaacgacgtcgtgtaaa 177

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	177	100.0	560	ADQ48575	Adq48575 OpIE2 pro
2	177	100.0	564	Aav62487	Aav62487 O. pseudo
3	177	100.0	2773	Aa161306	Aa161306 p22Op2F e
4	175.4	99.1	5038	ADQ48539	Adq48539 Viral vec
5	41	23.2	141	ADQ48576	Adq48576 Viral vec
6	32	18.1	921	Aaa02474	Aaa02474 Human col
7	30.8	17.4	1268	Abi18801	Abi18801 Drosophil
8	30.8	17.4	1342	Abi27503	Abi27503 Drosophil
9	30.8	17.4	3448	Abi27502	Abi27502 Drosophil
10	30.8	17.4	3600	Abi18800	Abi18800 Drosophil
11	30.8	17.4	29169	10 AAD64735	Aad64735 Human car
12	30.6	17.3	1404	ADT47065	Adt47065 Bacterial
13	30	16.9	657	ADT45632	Adt45632 Plant iso
14	30	16.9	1710	ABD14088	Abd14088 Pseudomon
15	30	16.9	3107	13 ADS89385	Ads89385 Oligonuc
16	30	16.9	3501	13 ADS89483	Ads89483 Oligonuc
17	30	16.9	6107	6 ABK31430	Abk31430 Signal tr
18	30	16.9	6107	6 ABK31430	Abk31430 Chemical
19	30	16.9	6107	6 ABK31430	Abk31430 Human gen
20	29.8	16.8	1370	6 ADF83430	Adf83430 Bread whe

21	29.8	16.8	1789	4 ABL19245	Abi19245 Drosophil
22	29.8	16.8	3915	4 ABL19244	Abi19244 Drosophil
23	29.6	16.7	300	3 AAA01209	Aaa01209 Human col
24	29.2	16.5	1593	13 ADT45695	Adt45695 Bacterial
25	28.6	16.2	585	11 ABD13766	Abd13766 Pseudomon
26	28.6	16.2	823	12 ACH89781	Ach89781 Human gen
27	28.6	16.2	915	5 AAS70780	Aas70780 DNA enco
28	28.6	16.2	1425	4 AAS51490	Aas51490 Pseudomon
29	28.6	16.2	1425	8 ACA19456	ACA19456 Prokaryot
30	28.6	16.2	1434	13 ADS56860	Ads56860 Bacterial
31	28.6	16.2	1443	11 ABD13493	Abd13493 Pseudomon
32	28.6	16.2	1551	11 ABD13719	Abd13719 Pseudomon
33	28.6	16.2	2085	11 ABD13633	Abd13633 Pseudomon
34	28.6	16.2	2091	2 AAX06989	Aax06989 Human neu
35	28.6	16.2	2091	6 ABK90730	Abk90730 cDNA enco
36	28.6	16.2	2502	2 AAX06988	Aax06988 Human neu
37	28.6	16.2	2502	6 ABK90731	Abk90731 cDNA enco
38	28.4	16.0	876	13 ADT43165	Adt43165 Bacterial
39	28.4	16.0	1320	8 ACA45469	ACA45469 Prokaryot
40	28.4	16.0	29040	11 ADL27170	Adl27170 Human gen
41	28.4	16.0	29322	9 ADA03092	Ada03092 Human LFN
42	28.4	16.0	29322	9 ADA66376	Ada66376 Human LFN
43	28.4	16.0	29322	10 ADB72830	Adb72830 Human LFN
44	28.4	16.0	75839	11 ACN43958	Acn43958 Human gen
45	28.2	15.9	283	3 AAC23731	Aac23731 Human sec

ALIGNMENTS

RESULT 1
ADQ48575
ID ADQ48575 standard; DNA; 560 BP.
XX
AC ADQ48575;
XX
DT 09-SEP-2004 (first entry)
XX
DE OpIE2 promoter DNA sequence.
XX
KW viral vector; recombination site; recombinant virus;
KW replication-defective particle generation; gene expression inhibition;
KW gene therapy vector; ds; OpIE2 promoter.
XX
OS Unidentified.
XX
PN WO2004009768-A2.
XX
PD 29-JAN-2004.
XX
PF 18-JUL-2003; 2003WO-US022437.
XX
PR 18-JUL-2002; 2002US-0396335P.
PR 26-JUL-2002; 2002US-039617P.
PR 19-NOV-2002; 2002US-0427231P.
PR 24-MAR-2003; 2003US-0456496P.
PR 03-JUN-2003; 2003US-0474940P.
XX
(INVI-) INVITROGEN CORP.
PA (BENN-) BENNETT R P.
PA (WELC/) WELCH P J.
PA (HARW/) HARWOOD S.
PA (MADD/) MADDEN K.
PA (FRIM/) FRIMPONG K.
(FRAN/) FRANK K E.
XX
Bennett RP, Welch PJ, Harwood S, Madden K, Frimpong K, Franke KE;
WPI; 2004-132944/13.
XX
Novel nucleic acid molecule comprising all or a portion of one or more
viral genome and further comprising two or more recombination sites that
do not substantially recombine with each other, useful as gene therapy.

XX PS Disclosure; Fig 16; 555pp; English.

XX CC The invention comprises a nucleic acid molecule consisting of all or a

XX CC portion of one or more viral/baculoviral genomes and further containing

XX CC at least two recombination sites that do not substantially recombine with

XX CC each other. One or more of the recombination sites is capable of

XX CC undergoing recombination with a compatible recombination site in the

XX CC presence of one or more proteins active in lambda recombination. The

XX CC nucleic acid of the invention replicates in prokaryotic and eukaryotic

XX CC cells. The nucleic acid of the invention is useful for constructing a

XX CC recombinant virus, generating replication-defective particles,

XX CC preventing/inhibiting the expression of one or more genes in an organism,

XX CC and are useful as gene therapy vectors. The nucleic acid of the invention

XX CC is also useful for producing and expressing fusion polypeptides. The

XX CC present DNA sequence represents the OpIE2 promoter that was used in the

XX CC exemplification of the invention.

SQ Sequence 560 BP; 144 A; 156 C; 116 G; 144 T; 0 U; 0 Other;

Query Match 100.0%; Score 177; DB 12; Length 560;

Best Local Similarity 100.0%; Pred. No. 2.6e-48;

Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTATCGTGACAGGAGCGCAGCTTCTCTGTGTTGCTTAACCGCAGCGCGCAACTCCT 60

DB 355 GTCTTATCGTGACAGGAGCGCAGCTTCTCTGTGTTGCTTAACCGCAGCGCGCAACTCCT 414

QY 61 TATCGGAACAGGAGCGCGCTCCATATCAGCGCGCGTTTCTCATGCGCGTGACCGGACA 120

DB 415 TATCGGAACAGGAGCGCGCTCCATATCAGCGCGCGTTTCTCATGCGCGTGACCGGACA 474

QY 121 CGAGGCGCGCGTCCCGCTTATCGCGCCTATAAATACAGCCCGCAACGATCTGGTAAA 177

DB 475 CGAGGCGCGCGTCCCGCTTATCGCGCCTATAAATACAGCCCGCAACGATCTGGTAAA 531

RESULT 2

ID AAV62487 standard; DNA; 564 BP.

XX AC AAV62487;

XX DT 17-OCT-2003 (revised)

XX DT 19-JAN-1999 (first entry)

XX DE O. pseudotsugata multicapsid nucleopolyhedrosis virus ie2 promoter.

XX KW Orgyia pseudotsugata; multicapsid; nucleopolyhedrosis virus; OpMNPV;

XX KW Op ie2; promoter; shuttle vector; transformation; melanotransferrin;

XX KW immediate early baculovirus promoter; prokaryotic; transcription;

XX KW bleomycin/bleomycin-type antibiotic; insect cell; transposon;

XX KW ion transport peptide hormone; ss.

XX OS Orgyia pseudotsugata; polyhedrosis virus.

XX PN WO9844141-A2.

XX PD 08-OCT-1998.

XX PF 26-MAR-1998; 98WO-CA000282.

XX PR 27-MAR-1997; 97US-0049946P.

XX PR 28-JAN-1998; 98CA-02221819.

XX XX (UYBR-) UNIV BRITISH COLUMBIA.

XX PI Grigliatti TA, Theilmann DA, Pfeifer TA, Hegedus DD;

XX WPI; 1998-557129/47.

XX DR Expression vectors for transforming insect cells from disparate lines -

XX PT useful to express heterologous DNA, e.g. to allow study of gene

XX PT

PT expression and produce commercially important proteins.

XX Claim 10; Page 82; 121pp; English.

XX CC This represents a Orgyia pseudotsugata multicapsid nucleopolyhedrosis

XX CC virus (OpMNPV) immediate early 2 (Op ie2) promoter sequence. The

XX CC invention provides a new shuttle vector for transforming insect cells

XX CC that comprises: (i) prokaryotic origin of replication; (ii) insect

XX CC promoter having homology to, and capable of functioning as, an immediate

XX CC early baculovirus promoter; (iii) prokaryotic promoter sequence, and (iv)

XX CC selectable marker capable of conferring resistance to a bleomycin/

XX CC phleomycin-type antibiotic under transcriptional control of (ii) and

XX CC (iii), in insect and prokaryotic cells respectively. The vectors can be

XX CC used to stably transform (especially insect) cells with heterologous DNA,

XX CC useful to allow study of gene expression and direct expression of

XX CC heterologous gene products e.g. commercially important proteins. They are

XX CC especially useful to allow expression of the heterologous

XX CC melanotransferrin, ion transport peptide hormones or biologically active

XX CC derivatives in insect cells. They enable transformation of insect cell

XX CC lines from disparate species, allowing screening of lines for optimum

XX CC post-translational modification of particular proteins. Shuttle vectors

XX CC further comprising DNA transposable elements defining a transposon can be

XX CC used to optimise heterologous protein expression and facilitate selection

XX CC of desired transformants. (Updated on 17-OCT-2003 to standardise OS

XX CC field)

XX SQ Sequence 564 BP; 147 A; 160 C; 116 G; 141 T; 0 U; 0 Other;

Query Match 100.0%; Score 177; DB 2; Length 564;

Best Local Similarity 100.0%; Pred. No. 2.6e-48;

Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTTATCGTGACAGGAGCGCAGCTTCTCTGTGTTGCTTAACCGCAGCGCGCAACTCCT 60

DB 351 GTCTTATCGTGACAGGAGCGCAGCTTCTCTGTGTTGCTTAACCGCAGCGCGCAACTCCT 410

QY 61 TATCGGAACAGGAGCGCGCTCCATATCAGCGCGCGTTTCTCATGCGCGTGACCGGACA 120

DB 411 TATCGGAACAGGAGCGCGCTCCATATCAGCGCGCGTTTCTCATGCGCGTGACCGGACA 470

QY 121 CGAGGCGCGCGTCCCGCTTATCGCGCCTATAAATACAGCCCGCAACGATCTGGTAAA 177

DB 471 CGAGGCGCGCGTCCCGCTTATCGCGCCTATAAATACAGCCCGCAACGATCTGGTAAA 527

RESULT 3

ID AAL61306 standard; DNA; 2773 BP.

XX AC AAL61306;

XX DT 22-SEP-2003 (first entry)

XX DE p22Op2F expression vector for insect cells.

XX KW Multimeric protein; interleukin 5; IL5; TNFalpha; inflammatory disease;

XX KW tumour necrosis factor alpha; gene therapy; arthritis; gene; ds.

XX OS Unidentified.

XX PN WO2003042244-A2.

XX PD 22-MAY-2003.

XX PR 15-NOV-2002; 2002WO-DK000764.

XX PR 16-NOV-2001; 2001DK-00001702.

XX PR 16-NOV-2001; 2001US-0331575P.

XX XX (PHAR-) PHARMEXA AS.

XX XX (KLYS/) KLYSNER S.

XX XX (NIEL/) NIELSEN F S.

XX XX (BRAT/) BRATT T.

PA (VOLD/) VOLDORGB B.
PA (MOUR/) MOURITSEN S.
XX
PI Klynsen S, Nielsen FS, Bratt T, Voldborg B, Mouritsen S;
XX
XX WPI; 2003-449558/42.
DR
XX
XX New immunogenic analogue of a polymeric protein, useful for preparing a
PT composition for treating inflammatory diseases e.g. arthritis.
PT
XX
XX Disclosure; Page 195-196; 196pp; English.
XX
XX The invention relates to immunogenic analogues of multimeric proteins
CC such as immunogenic variants of interleukin 5 (IL5) and tumour necrosis
CC factor alpha (TNF, TNFalpha) and methods for production of immunogenic
CC analogues. The immunogenic analogue is useful for preparing a composition
CC for treating inflammatory diseases, e.g., arthritis. It is also used in
CC gene therapy. The present sequence is p220p2F expression vector for
CC insect cells. This sequence is used to illustrate the method of the
CC invention
XX
SQ Sequence 2773 BP; 669 A; 750 C; 673 G; 681 T; 0 U; 0 Other;
Query Match 100.0%; Score 177; DB 9; Length 2773;
Best Local Similarity 100.0%; Pred. No. 3.8e-48;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCTATCGTGACAGAGCCAGCTTCCTGTTGCTTAACCGCAGCGGACGCAACTCTCT 60
DB 355 GTCCTATCGTGACAGAGCCAGCTTCCTGTTGCTTAACCGCAGCGGACGCAACTCTCT 414
QY 61 TATCGGAACAGGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGGTGACCGGACA 120
DB 415 TATCGGAACAGGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGGTGACCGGACA 474
QY 121 CGAGGCGCCGTCGCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 177
DB 475 CGAGGCGCCGTCGCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 531
RESULT 4
ADQ48539
ID ADQ48539 standard; DNA; 5038 BP.
XX
XX
XX ADQ48539;
XX
DT 09-SEP-2004 (first entry)
XX
DE Viral vector-related plasmid - pIB/V5-His-DEST.
XX
XX viral vector; recombination site; recombinant virus;
KW replication-defective particle generation; gene expression inhibition;
KW gene therapy vector; ds; plasmid.
XX
XX Unidentified.
OS
XX
XX WO2004009768-A2.
XX
XX 29-JAN-2004.
XX
PF 18-JUL-2003; 2003WO-US022437.
XX
XX 18-JUL-2002; 2002US-0396335P.
PR 26-JUL-2002; 2002US-0398617P.
PR 19-NOV-2002; 2002US-0427231P.
PR 24-MAR-2003; 2003US-0456496P.
PR 03-JUN-2003; 2003US-0474940P.
XX
XX (INVI-) INVITROGEN CORP.
PA (BENN/) BENNETT R P.
PA (WELC/) WELCH P J.
PA (HARW/) HARWOOD S.
PA (MADD/) MADDEN K.

PA (FRIM/) FRIMPONG K.
PA (FRAN/) FRANK K E.
XX
PI Bennett RP, Welch PJ, Harwood S, Madden K, Frimpong K, Franke KE;
XX
XX WPI; 2004-132944/13.
DR
XX
XX Novel nucleic acid molecule comprising all or a portion of one or more
PT viral genome and further comprising two or more recombination sites that
PT do not substantially recombine with each other, useful as gene therapy.
XX
XX Example 18; Page 395-403; 555pp; English.
XX
XX The invention comprises a nucleic acid molecule consisting of all or a
CC portion of one or more viral/baculoviral genomes and further containing
CC at least two recombination sites that do not substantially recombine with
CC each other. One or more of the recombination sites is capable of
CC undergoing recombination with a compatible recombination site in the
CC presence of one or more proteins active in lambda recombination. The
CC nucleic acid of the invention replicates in prokaryotic and eukaryotic
CC cells. The nucleic acid of the invention is useful for constructing a
CC recombinant virus, generating replication-defective particles,
CC preventing/inhibiting the expression of one or more genes in an organism,
CC and are useful as gene therapy vectors. The nucleic acid of the invention
CC is also useful for producing and expressing fusion polypeptides. The
CC present DNA sequence represents a plasmid that was used in the
CC exemplification of the invention.
XX
SQ Sequence 5038 BP; 1342 A; 1178 C; 1198 G; 1320 T; 0 U; 0 Other;
Query Match 99.1%; Score 175.4; DB 12; Length 5038;
Best Local Similarity 99.4%; Pred. No. 1.5e-47;
Matches 176; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCCTATCGTGACAGGACGCGCTTCCTGTTGCTTAACCGCAGCGGACGCAACTCTCT 60
DB 351 GTCCTATCGTGACAGGACGCGCTTCCTGTTGCTTAACCGCAGCGGACGCAACTCTCT 410
QY 61 TATCGGAACAGGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGGTGACCGGACA 120
DB 411 TATCGGAACAGGACGCGCTCCATATCAGCCGCGGTTATCTCATGCGGTGACCGGACA 470
QY 121 CGAGGCGCCGTCGCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 177
DB 471 CGAGGCGCCGTCGCGCTTATCGCGCTTATAAATACAGCCGCAACGATCTGGTAAA 527
RESULT 5
ADQ48576
ID ADQ48576 standard; DNA; 141 BP.
XX
XX ADQ48576;
XX
DT 09-SEP-2004 (first entry)
XX
XX
XX Viral vector-related plasmid pIB/V5-His-DEST recombination region #1.
DE
XX
XX viral vector; recombination site; recombinant virus;
KW replication-defective particle generation; gene expression inhibition;
KW gene therapy vector; ds; plasmid; recombination region.
XX
XX Unidentified.
OS
XX
XX WO2004009768-A2.
XX
XX 29-JAN-2004.
XX
XX 18-JUL-2003; 2003WO-US022437.
XX
XX 18-JUL-2002; 2002US-0396335P.
PR 26-JUL-2002; 2002US-0398617P.
PR 19-NOV-2002; 2002US-0427231P.
PR 24-MAR-2003; 2003US-0456496P.
PR 03-JUN-2003; 2003US-0474940P.

```
PR 03-JUN-2003; 2003US-0474940P.
XX
PA (INVI-) INVITROGEN CORP.
PA (BEN)/ BENNETT R P.
PA (WELC)/ WELCH P J.
PA (HARW)/ HARWOOD S.
PA (MADD)/ MADDEN K.
PA (FRIM)/ FRIMPONG K.
PA (FRAN)/ FRANK K E.
XX
XX Bennett RP, Welch PJ, Harwood S, Madden K, Frimpong K, Franke KE;
XX WPI; 2004-132944/13.
XX
XX Novel nucleic acid molecule comprising all or a portion of one or more
XX viral genome and further comprising two or more recombination sites that
XX do not substantially recombine with each other, useful as gene therapy.
XX
XX Disclosure; Fig 17; 555pp; English.
XX
XX The invention comprises a nucleic acid molecule consisting of all or a
XX portion of one or more viral/baculoviral genomes and further containing
XX at least two recombination sites that do not substantially recombine with
XX each other. One or more of the recombination sites is capable of
XX undergoing recombination with a compatible recombination site in the
XX presence of one or more proteins active in lambda recombination. The
XX nucleic acid of the invention replicates in prokaryotic and eukaryotic
XX cells. The nucleic acid of the invention is useful for constructing a
XX recombinant virus, generating replication-defective particles,
XX preventing/inhibiting the expression of one or more genes in an organism,
XX and are useful as gene therapy vectors. The nucleic acid of the invention
XX is also useful for producing and expressing fusion polypeptides. The
XX present DNA sequence represents a recombination region of a plasmid that
XX was used in the exemplification of the invention.
XX
XX Sequence 141 BP; 43 A; 36 C; 27 G; 34 T; 0 U; 1 Other;
XX
XX Query Match 23.2%; Score 41; DB 12; Length 141;
XX Best Local Similarity 100.0%; Pred. No. 0.0014;
XX Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX Qy 137 CTTATCGCGCCTATAATACAGCCGCAACGATCTGGTAA 177
XX |||||
XX 1 CTTATCGCGCCTATAATACAGCCGCAACGATCTGGTAA 41
XX
XX
XX RESULT 6
XX ID AAA02474/c
XX AC AAA02474;
XX
XX 19-MAY-2000 (first entry)
XX
XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:2465.
XX
XX Human; colon cancer; tumour; diagnosis; gene expression product; probe;
XX detection; cancerous state; metastasis; identification; breast cancer;
XX oestrogen receptor-positive breast cancer; therapy;
XX oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
XX
XX WO9958675-A2.
XX
XX 18-NOV-1999.
XX
XX 13-MAY-1999; 99WO-US010602.
XX
XX 14-MAY-1998; 98US-0085426P.
XX 15-MAY-1998; 98US-0085537P.
XX 15-MAY-1998; 98US-0085696P.
XX 21-OCT-1998; 98US-0105234P.
XX
```

```
PR 27-OCT-1998; 98US-0105877P.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
XX Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.
XX
XX Polynucleotide library used to determine cancerous states of mammalian
XX cells.
XX
XX Claim 1; Page 989; 1097pp; English.
XX
XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
XX libraries constructed from human colon cancer cell lines. The present
XX invention also describes a method of detecting differentially expressed
XX genes correlated with a cancerous state of a mammalian cell, comprising
XX detecting at least one differentially expressed gene product in a test
XX sample derived from a cell suspected of being cancerous, where detection
XX of the differentially expressed gene product is correlated with a
XX cancerous state of the cell from which the test sample was derived. The
XX polynucleotide sequences can be used in a method for detecting
XX differentially expressed genes correlated with a cancerous state of a
XX mammalian cell. The polynucleotides can also be used as probes for
XX detecting and mapping related genes. They can be used in diagnosis and
XX prognosis of diseases and disorders (e.g. identification of pre-
XX metastatic or metastatic cancerous states, stages of cancer, or
XX responsiveness of cancer to therapy). This is particularly for breast
XX cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
XX negative breast cancer, lung cancer, and colon cancer
XX
XX Sequence 921 BP; 239 A; 214 C; 205 G; 223 T; 0 U; 40 Other;
XX
XX Query Match 18.1%; Score 32; DB 3; Length 921;
XX Best Local Similarity 51.5%; Pred. No. 2.1;
XX Matches 68; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
XX
XX Qy 16 GAGCCAGCTTCTGTGTGCTTAACCGCAGCGGACGCAACTCTTATCGGAACAGGACG 75
XX |||||
XX 157 GGCAGCTGCGCNCAGGTGGCTGTCCACGCGGTCTCCGCGCTGCCGTCGCGGGT 98
XX |||||
XX 76 GCGCTCCATATCAGCCGCGGTTATCTCATCGCGGTACCGGACACAGAGGCCCGTCCC 135
XX |||||
XX 97 GCGCTCNAANCCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 38
XX |||||
XX 136 GCTTATCGCGCC 147
XX |||||
XX 37 TAAGATCGNGCC 26
XX
XX
XX RESULT 7
XX ABL18801
XX ID ABL18801 standard; DNA; 1268 BP.
XX
XX AC ABL18801;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 7876.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
```


CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 3448 BP; 894 A; 817 C; 801 G; 936 T; 0 U; 0 Other;

Query Match 17.4%; Score 30.8; DB 4; Length 3448;
Best Local Similarity 55.7%; Pred. No. 7.1;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Oy 17 ACGCCAGCTTCCTGTGTTGCTAACCGGACCGGAGCGCAACTCCTTATCGGAACAGGACGC 76
Db 1667 ACAACAGCTTCGGGTGGAGGTGCCCAAGCGGTCTCTAGTCTTGGCGGTCCGGTGGC 1608

Oy 77 GCCTCCATATCAGCGCGCGTATCTCATGCGCGTGACCGGACACG 122
Db 1607 GTTGGCACATTGGCCATACAAATCCTCAAGTCCAGAAAGTACAAG 1562

RESULT 10
ABL18800/c
ID ABL18800 standard; DNA; 3600 BP.
XX ABL18800;
AC ABL18800;
XX
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 7873.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 7873; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 3600 BP; 946 A; 856 C; 823 G; 975 T; 0 U; 0 Other;

Query Match 17.4%; Score 30.8; DB 4; Length 3600;
Best Local Similarity 55.7%; Pred. No. 7.2;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Oy 17 ACGCCAGCTTCCTGTGTTGCTAACCGGACCGGAGCGCAACTCCTTATCGGAACAGGACGC 76
Db 1667 ACAACAGCTTCGGGTGGAGGTGCCCAAGCGGTCTCTAGTCTTGGCGGTCCGGTGGC 1608

Oy 77 GCCTCCATATCAGCGCGCGTATCTCATGCGCGTGACCGGACACG 122
Db 1607 GTTGGCACATTGGCCATACAAATCCTCAAGTCCAGAAAGTACAAG 1562

RESULT 11
AAD64735
ID AAD64735 standard; DNA; 29169 BP.
XX AAD64735;
AC AAD64735;
XX
XX 11-MAR-2004 (first entry)
XX
XX Human carcinoma-associated (CA) gene TBX21.
XX
XX Carcinoma-associated protein; CAP; acinar cell carcinoma; fibrosarcoma;
XX Kaposi's sarcoma; breast cancer; Hairy cell leukaemia; human; ds.
XX
XX Homo sapiens.
XX
XX US2003099963-A1.
XX
XX 29-MAY-2003.
XX
XX 20-MAR-2002; 2002US-00105613.
XX
XX 22-DEC-2000; 2000US-00747377.
XX 02-MAR-2001; 2001US-00798586.
XX 08-NOV-2001; 2001US-00052482.
XX
XX (MORR/) MORRIS D W.
XX (ENGE/) ENGELHARD E K.
XX
XX Morris DW, Engelhard EK;
XX
XX WPI; 2003-874605/81.
XX
XX Novel recombinant carcinoma-associated protein such as mouse or human
XX TBX21 protein, useful for screening a bioactive agent capable of binding
XX to carcinoma associated protein, and for evaluating the effect of a
XX candidate carcinoma drug.
XX
XX Disclosure; SEQ ID NO 4; 0pp; English.
XX
XX The present invention relates to novel recombinant carcinoma-associated
XX protein (CAP) useful for screening a bioactive agent capable of binding
XX to carcinoma associated (CA) protein and for evaluating the effect of a
XX candidate carcinoma drug. The invention also describes the use of novel
XX compositions for use in screening methods and provides compositions and
XX methods associated with altered expression of TBX21 in cancer. Suitable
XX cancers which can be diagnosed or screened by the invention includes
XX acinar cell carcinoma, fibrosarcoma, Kaposi's sarcoma, breast cancer and
XX Hairy cell leukaemia. The present sequence is human carcinoma-associated
XX (CA) gene
XX
SQ Sequence 29169 BP; 5725 A; 7553 C; 7224 G; 8641 T; 0 U; 26 Other;

Query Match 17.4%; Score 30.8; DB 10; Length 29169;
Best Local Similarity 61.0%; Pred. No. 12;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Oy 11 GACAGGAGCGGAGCTTCTGTGTTGCTAACCGGACCGGAGCGCAACTCCTTATCGGAACA 70
Db 8982 GACACGACGCGCTGCTGTGTGGAACTGGAGGGAGGAGCCCTTATCTCCGG 9041

Oy 71 GGACGCGCCTTCATATCAGCGC 92
Db 9042 CCCCTGCGCCACCTCCCGC 9063

Query Match	17.3%	Score 30.6;	DB 13;	Length 1404;
Best Local Similarity	62.3%;	Pred. No. 6.7;		
Matches	48;	Conservative	0;	Mismatches 29; Indels 0; Gaps 0;
QY	99	ATCTCATGCGCGTGACCGGACACGAGCGCGCGCTTATCGCGCTATAAATACAG	158	
Db	464	ACCGCAAGGCGTGAACGTGCCGATGTGGTGTGCGCGCTTTTCGGCCAAGG	523	
QY	159	CCCGCAACGATCTGGTA	175	
Db	524	ACCGCGACGATCTGGAA	540	
RESULT 13				
ADI45632/c				
ID	ADI45632	standard; DNA; 657 BP.		
XX	ADI45632;			
XX	22-APR-2004	(first entry)		
XX	Plant isoprenoid biosynthesis-associated DNA #6.			
XX	isoprenoid biosynthesis; ss; isopentenyl diphosphate; IPP;			
KW	dimethylallyl alcohol; DMAPP; short-chain plastid prenyltransferase;			
KW	gibberellin; carotenoid; abscisic acid; tocopherol; plastoquinone;			
KW	phyllolignone; mevalonate pathway; phytosterol; brassinosteroid;			
KW	ubiquinone; monoterpene; sesquiterpene; protein prenylation; chlorophyll;			
KW	haeme; yield.			
XX	Unidentified.			
OS	US2004010815-A1.			
PN	15-JAN-2004.			
PD	26-SEP-2002; 2002US-00259194.			
XX	26-SEP-2001; 2001US-0325277P.			
PR	04-APR-2002; 2002US-0370620P.			
PR	04-APR-2002; 2002US-0370743P.			
XX	(LANG/) LANGE B M.			
PA	(GHAS/) GHASSEMIAN M.			
PA	(BRIG/) BRIGGS S P.			
PA	(COOP/) COOPER B.			
PA	(GLAZ/) GLAZEBROOK J.			
PA	(GOFF/) GOFF S A.			
PA	(KATA/) KATAGIRI F.			
PA	(KREP/) KREPS J.			
PA	(MOUG/) MOUGHAMER T.			
PA	(PROV/) PROVART N.			
PA	(RICK/) RICHE D.			
PA	(ZHUT/) ZHU T.			
XX	Lange BM, Ghassemian M, Briggs SP, Cooper B, Glazebrook J;			
PI	Goff SA, Katagiri F, Kreps J, Moughamer T, Provart N, Riche D;			
PI	Zhu T;			
XX	WPI; 2004-090562/09.			
XX	New isolated polynucleotides and polypeptides associated with isoprenoid			
PT	synthesis in plants, useful for producing transgenic plants, for targeted			
PT	gene disruption, as well as markers or probes.			
XX	Disclosure; SEQ ID NO 563; 117pb; English.			
XX	The invention relates to a polynucleotide (or its complement, protein			
CC	encoding fragment or reverse complement), comprising a nucleotide			
CC	sequence encoding a polypeptide comprising an amino acid sequence			
CC	involved in or associated with the biosynthesis of isoprenoids in a rice			
CC	plant. Also included are an isolated polypeptide involved in or			
CC	associated with the biosynthesis of isoprenoids in a plant an expression			

CC cassette comprising the polynucleotide, a host cell comprising the
CC expression cassette, and a transgenic plant comprising the expression
CC cassette. The polypeptides and polynucleotides include those associated
CC with the biosynthesis of isopentenyl diphosphate (IPP) and dimethylallyl
CC pyrophosphate (DMAPP), the biosynthesis of short-chain plastid
CC prenyltransferases, the biosynthesis of gibberellins, the biosynthesis of
CC carotenoids and/or abscisic acids, the biosynthesis of tocopherols,
CC plastoquinone and/or phyloquinone biosynthesis, the mevalonate pathway,
CC phytosterol and brassinosteroid metabolism, biosynthesis of ubiquinone,
CC biosynthesis of monoterpenes and sesquiterpenes, protein prenylation, and
CC biosynthesis of chlorophyll or haeme. Also disclosed are banana, wheat
CC and corn homologues of some of the rice polynucleotides. The
CC polynucleotides are useful for producing transgenic plants, where the
CC genome is augmented by a nucleic acid molecule of the invention, or in
CC which the corresponding gene has been disrupted, e.g. to result in a
CC loss, a decrease or an alteration in the function of the product encoded
CC by the gene. The plants may also have increased yields and/or produce a
CC better quality product than the corresponding wild-type plant. The
CC nucleic acid molecules are useful for targeted gene disruption, as well
CC as markers and probes. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20040010815. The present sequence
CC is an isoprenoid biosynthesis- associated DNA included in the sequence
CC listing but not mentioned anywhere else in the specification.

XX
SQ Sequence 657 BP; 109 A; 257 C; 178 G; 113 T; 0 U; 0 Other;

Query Match 16.9%; Score 30; DB 12; Length 657;
Best Local Similarity 53.4%; Pred. No. 8.9;
Matches 63; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 29 TGTGTGCTACCGCGGACGCACTCTTATCGAAGACGAGCGGCTCCATATCA 88
Db 506 TGTGTGCTACCGCGGACGCACTCTTATCGAAGACGAGCGGCTCCATATCA 88
Qy 89 GCCGCGCTTATCTCATCGGCTGACCGACACGAGGCGCGCTCCGCTTATCGCGC 146
Db 446 GCCGCGATTTGCGGTCTGTCGCGGCTCTGACGCGGCGCGCTCCGCGC 389

RESULT 14
ABD14088/C
ID ABD14088 standard; DNA; 1710 BP.
XX
AC ABD14088;
XX
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #12692.
XX
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO80517.
XX
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,

PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 12692; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html

XX
SQ Sequence 1710 BP; 380 A; 455 C; 526 G; 349 T; 0 U; 0 Other;

Query Match 16.9%; Score 30; DB 11; Length 1710;
Best Local Similarity 57.4%; Pred. No. 11;
Matches 54; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 81 CCATATCAGCGCGGCTTATCTCATGCGGTGACGCGACGAGCGCGCTCCGCTTA 140
Db 414 CCAAAATAGCGGCTCGATCTCTCGGCGAAGTACGCGGCGACCTGGCGAGCGCATTTGCCCC 355
Qy 141 TCGCGCTTATAAATACAGCGCGCAACGATCTGGT 174
Db 354 TCGCGCGTCAACTCTCGCATTCACCTGCGAGGT 321

RESULT 15
ADS89385/C
ID ADS89385 standard; DNA; 3107 BP.

XX
AC ADS89385;
XX
XX
DT 18-NOV-2004 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:401.
XX
KW ss; cell proliferative disorder; breast; methylation; cytostatic;
KW gene therapy; single nucleotide polymorphism; SNP.
XX
OS Unidentified.

XX
PN WO2004035803-A2.
XX
PD 29-APR-2004.
XX
PF 01-OCT-2003; 2003WO-EP010881.
XX
PR 01-OCT-2002; 2002DE-01045779.
PR 07-JAN-2003; 2003DE-01000096.
PR 17-APR-2003; 2003DE-01017955.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Foekens J, Harbeck N, Koenig T, Maier S, Martens J, Model P;
PI Nimmrich I, Rujaan T, Schmitt A, Schmitt M, Look MP, Marx A;
XX
DR WPI; 2004-348468/32.
XX

PT Predicting responsiveness of a subject with breast cell proliferative
PT disorder, useful for treating or differentiating breast cell
PT proliferative disorders comprises analyzing methylation pattern of a
PT genomic DNA from the subject.

XX	Claim 25; SEQ ID NO 401; 104pp: English.
PS	The invention relates to a novel method for predicting the responsiveness
XX	of a subject with a cell proliferative disorder of the breast tissues to
CC	a therapy comprising analysing the methylation pattern of a target
CC	nucleic acid by contacting at least one of the target nucleic acids in a
CC	biological sample obtained from the subject prior to or during treatment.
CC	The method of the invention has cytostatic activity, and may have a use
CC	in gene therapy. The set of oligonucleotides comprising at least two of
CC	the oligomers are useful for detecting the cytosine methylation state
CC	and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC	methods, nucleic acid, oligonucleotide, and kit are useful for the
CC	treatment, characterisation, classification and/or differentiation, of
CC	breast cell proliferative disorders. The method is also useful for
CC	predicting the responsiveness of a subject with a cell proliferative
CC	disorder of the breast tissues to a therapy. The present sequence is used
CC	in the exemplification of the invention.
XX	
SQ	Sequence 3107 BP; 823 A; 120 C; 814 G; 1350 T; 0 U; 0 Other;
	Query Match 16.9%; Score 30; DB 13; Length 3107;
	Best Local Similarity 50.7%; Pred. No. 13;
	Matches 72; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY	6 ATCGTGACAGGACGCCAGCTTCCTGTGTTGTATACCGCAGCGGACGCAACTCTTATCG 65
DB	2079 AACCGGAAACACCCGGTCTCTCTCGGTCTCCAACCGCGCGCGGAATAACGCTCAACGCC 2020
QY	66 GAACGAGCACGGCCCTCCATATCAGCGCGCGGTATTCTCATCGCGGTGACCGGACACGAGG 125
DB	2019 GAACCCGGACTCTTCTCTCATTTTAACAACACTTCCTTAACGCGCCCCGAAACAGACCCG 1960
QY	126 CGCCCGTCCCGCTTATCGGCC 147
DB	1959 CACTTCTCTCCCAATATCGCGC 1938

Search completed: October 24, 2005, 18:58:05
Job time : 199.335 secs

Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 20:18:24 ; Search time 361.161 Seconds
(without alignments)
4044.488 Million cell updates/sec

Title: US-09-896-888A-1_COPY_351_527

Perfect score: 177

Sequence: 1 gctctatgtagcagcgc.....gcccgcaacgacatcgtgtaaa 177

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9772363 seqs, 4126298632 residues

Total number of hits satisfying chosen parameters: 19544726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubnpa/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubnpa/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubnpa/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubnpa/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubnpa/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubnpa/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubnpa/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubnpa/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubnpa/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubnpa/US10D_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubnpa/US10E_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/1/pubnpa/US10F_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/1/pubnpa/US10G_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/1/pubnpa/US10H_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/1/pubnpa/US10I_PUBCOMB.seq.*
- 23: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq.*
- 24: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/1/pubnpa/US11A_PUBCOMB.seq.*
- 26: /cgn2_6/ptodata/1/pubnpa/US11_NEW_PUB.seq.*
- 27: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq.*
- 28: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	177	100.0	462	9	US-09-896-888A-14
2	177	100.0	560	21	US-10-622-088-126
3	177	100.0	564	9	US-09-896-888A-1
4	177	100.0	2773	17	US-10-295-074-60
5	177	100.0	2773	21	US-10-846-911-60

6	177	100.0	2773	24	US-10-939-107-60	Sequence 60, Appl
7	175.4	99.1	5038	21	US-10-622-088-89	Sequence 89, Appl
8	41	23.2	147	21	US-10-622-088-127	Sequence 127, Appl
9	41	23.2	325	21	US-10-622-088-149	Sequence 149, Appl
10	32.2	18.2	1138	16	US-10-156-761-2538	Sequence 2538, Ap
11	32.2	18.2	9025608	16	US-10-156-761-1	Sequence 1, Appli
12	32	18.1	921	24	US-10-779-543-8561	Sequence 8561, Ap
13	30.8	17.4	1268	26	US-11-097-143-25442	Sequence 25442, A
14	30.8	17.4	1342	26	US-11-097-143-38495	Sequence 38495, A
15	30.8	17.4	3448	26	US-11-097-143-38494	Sequence 38494, A
16	30.8	17.4	3600	26	US-11-097-143-25441	Sequence 25441, A
17	30.6	17.3	1404	18	US-10-369-493-45503	Sequence 45503, A
18	30.2	17.1	1116	20	US-10-437-963-42126	Sequence 42126, A
19	30.2	17.1	2472	20	US-10-437-963-55473	Sequence 55473, A
20	30	16.9	657	18	US-10-259-194A-563	Sequence 563, App
21	30	16.9	6107	18	US-10-221-613-303	Sequence 303, App
22	29.8	16.8	1370	19	US-10-275-311A-10	Sequence 10, Appl
23	29.8	16.8	1789	26	US-11-097-143-26108	Sequence 26108, A
24	29.6	16.7	3915	26	US-11-097-143-26107	Sequence 26107, A
25	29.6	16.7	300	24	US-10-779-543-7296	Sequence 7296, Ap
26	29.6	16.7	1798	21	US-10-739-930-4094	Sequence 4094, Ap
27	29.6	16.7	3498	20	US-10-437-963-72869	Sequence 72869, A
28	29.4	16.6	1923	20	US-10-437-963-50761	Sequence 50761, A
29	29.4	16.6	5118	20	US-10-437-963-35867	Sequence 35867, A
30	29.2	16.5	1593	18	US-10-369-493-44133	Sequence 44133, A
31	28.8	16.3	646	13	US-09-925-065A-711960	Sequence 711960,
32	28.6	16.2	648	21	US-10-425-115-65984	Sequence 65984, A
33	28.6	16.2	823	17	US-10-029-386-22976	Sequence 22976, A
34	28.6	16.2	915	24	US-10-450-763-6584	Sequence 6584, Ap
35	28.6	16.2	1425	9	US-09-815-242-4072	Sequence 4072, Ap
36	28.6	16.2	1425	18	US-10-282-122A-7326	Sequence 7326, Ap
37	28.6	16.2	1434	18	US-10-369-493-32534	Sequence 32534, A
38	28.6	16.2	2091	9	US-09-817-647-22	Sequence 22, Appl
39	28.6	16.2	2091	9	US-09-877-665-22	Sequence 22, Appl
40	28.6	16.2	2091	14	US-10-136-573A-22	Sequence 22, Appl
41	28.6	16.2	2091	15	US-10-215-862-22	Sequence 22, Appl
42	28.6	16.2	2091	22	US-10-944-116-22	Sequence 22, Appl
43	28.6	16.2	2091	26	US-11-035-787-22	Sequence 22, Appl
44	28.6	16.2	2502	9	US-09-817-647-5	Sequence 5, Appli
45	28.6	16.2	2502	9	US-09-877-665-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-896-888A-14
; Sequence 14, Application US/09896888A
; Patent No. US20020116723A1
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Insect Expression Vectors
; FILE REFERENCE: 80021-44
; CURRENT APPLICATION NUMBER: US/09/896,888A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US/09/048,911
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/049,946
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Promoter
; OTHER INFORMATION: sequence of the OpMNPV ie2 gene
US-09-896-888A-14

Query Match 100.0%; Score 177; DB 9; Length 462;
Best Local Similarity 100.0%; Fred. No. 2e-54;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCTTATCGTGACAGGACGCCAGCTTCTGTGTTGCTTAACCGACGCCGAGCGCAACTCCT 60
Db 250 GTCTTATCGTGACAGGACGCCAGCTTCTGTGTTGCTTAACCGACGCCGAGCGCAACTCCT 309
Qy 61 TATCGGAACAGGACGCCGCTCCATATACAGCGCGGTTTATCTCATGCGCGTGACCGGACA 120
Db 310 TATCGGAACAGGACGCCGCTCCATATACAGCGCGGTTTATCTCATGCGCGTGACCGGACA 369
Qy 121 CGAGCGCCCGTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAA 177
Db 370 CGAGCGCCCGTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAA 426

RESULT 2

US-10-622-088-126
; Sequence 126, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth
; APPLICANT: Franke, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622.088
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/22437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 126
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: OpIE2 promoter sequence
US-10-622-088-126

Query Match 100.0%; Score 177; DB 21; Length 560;
Best Local Similarity 100.0%; Pred. No. 2e-54;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCTTATCGTGACAGGACGCCAGCTTCTGTGTTGCTTAACCGACGCCGAGCGCAACTCCT 60
Db 355 GTCTTATCGTGACAGGACGCCAGCTTCTGTGTTGCTTAACCGACGCCGAGCGCAACTCCT 414
Qy 61 TATCGGAACAGGACGCCGCTCCATATACAGCGCGGTTTATCTCATGCGCGTGACCGGACA 120
Db 415 TATCGGAACAGGACGCCGCTCCATATACAGCGCGGTTTATCTCATGCGCGTGACCGGACA 474
Qy 121 CGAGCGCCCGTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAA 177
Db 475 CGAGCGCCCGTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAA 531

RESULT 3

US-09-896-888A-1
; Sequence 1, Application US/09896888A
; Patent No. US20020116723A1
; GENERAL INFORMATION:

; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Insect Expression Vectors
; FILE REFERENCE: 80021-44
; CURRENT APPLICATION NUMBER: US/09/896,888A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US/09/048,911
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/049,946
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Orgyia pseudotsugata
US-09-896-888A-1

Query Match 100.0%; Score 177; DB 9; Length 564;
Best Local Similarity 100.0%; Pred. No. 2e-54;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCTTATCGTGACAGGACGCCAGCTTCTGTGTTGCTTAACCGACGCCGAGCGCAACTCCT 60
Db 351 GTCTTATCGTGACAGGACGCCAGCTTCTGTGTTGCTTAACCGACGCCGAGCGCAACTCCT 410
Qy 61 TATCGGAACAGGACGCCGCTCCATATACAGCGCGGTTTATCTCATGCGCGTGACCGGACA 120
Db 411 TATCGGAACAGGACGCCGCTCCATATACAGCGCGGTTTATCTCATGCGCGTGACCGGACA 470
Qy 121 CGAGCGCCCGTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAA 177
Db 471 CGAGCGCCCGTCCCGCTTATCGCGCTTATAAATACAGCCCGCAACGATCTGGTAAA 527

RESULT 4

US-10-295-074-60
; Sequence 60, Application US/10295074
; Publication No. US20030185845A1
; GENERAL INFORMATION:
; APPLICANT: Pharmexa A/S
; TITLE OF INVENTION: NOVEL IMMUNOGENIC MIMETICS OF MULTIMER PROTEINS
; FILE REFERENCE: P1013DK00
; CURRENT APPLICATION NUMBER: US/10/295,074
; CURRENT FILING DATE: 2002-11-15
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60
; LENGTH: 2773
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: p2Op2F expression vector for insect cells
; NAME/KEY: misc_recomb
; LOCATION: (561)..(566)
; OTHER INFORMATION: HindIII site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (573)..(578)
; OTHER INFORMATION: Aval site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (586)..(591)
; OTHER INFORMATION: EcoRI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (593)..(598)
; OTHER INFORMATION: BamHI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (625)..(630)
; OTHER INFORMATION: ClaI site
; FEATURE:

```

; OTHER INFORMATION: p22Op2F expression vector for insect cells
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (561)..(566)
; OTHER INFORMATION: HindIII site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (573)..(578)
; OTHER INFORMATION: Aval site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (586)..(591)
; OTHER INFORMATION: EcoRI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (593)..(598)
; OTHER INFORMATION: BamHI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (625)..(630)
; OTHER INFORMATION: ClaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (629)..(634)
; OTHER INFORMATION: ClaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (1156)..(1161)
; OTHER INFORMATION: ApaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2128)..(2133)
; OTHER INFORMATION: PstI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2204)..(2209)
; OTHER INFORMATION: NcoI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2284)..(2289)
; OTHER INFORMATION: Aval site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2294)..(2299)
; OTHER INFORMATION: Aval, SmaI, and XmaI site
; FEATURE:
; NAME/KEY: misc_recomb
; LOCATION: (2551)..(2556)
; OTHER INFORMATION: ApaI site
; US-10-846-911-60

```

```

Query Match      100.0%; Score 177; DB 21; Length 2773;
Best Local Similarity 100.0%; Pred. No. 2.6e-54;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    1   GTCTTATCGTGACAGGACGCCAGCTTCCTGTGCTGTAAACGCAGCCGGACGCAACTCCT   60
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    355  GTC TTATCGTGACAGGACGCCAGCTTCCTGTGCTGTAAACGCAGCCGGACGCAACTCCT   414

Qy    61  TATCGGAACAGGACGGCGCTCCATATCATGCGCGGCTTATCTCATGCGCGTAGCCGGACA   120
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    415  TATCGGAACAGGACGGCGCTCCATATCATGCGCGGCTTATCTCATGCGCGTAGCCGGACA   474

Qy    121 CGAGGGCGCCGTCGCGCTTATCGCGGCTATAAATACAGCCCCGACGATCTGGTAAA   177
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    475  CGAGGGCGCCGTCGCGCTTATCGCGGCTATAAATACAGCCCCGACGATCTGGTAAA   531

RESULT 6
US-10-939-107-60
; Sequence 60, Application US/10939107
; Publication No. US20050180947A1
GENERAL INFORMATION:
```

APPLICANT: Pedersen, Hans Rudolf
APPLICANT: Ebert, Bjarke
APPLICANT: Pedersen, Louise Henriette
APPLICANT: Rasmussen, Peter Birk
TITLE OF INVENTION: Novel Application of Vaccination Against TNF-alpha
FILE REFERENCE: 674542-2020
CURRENT APPLICATION NUMBER: US/10/939,107
CURRENT FILING DATE: 2004-09-10
PRIOR APPLICATION NUMBER: PCT/DK03/00147
PRIOR FILING DATE: 2003-03-11
PRIOR APPLICATION NUMBER: 60/363,128
PRIOR FILING DATE: 2002-03-11
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn version 3.1
SEQ ID NO 60
LENGTH: 2773
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: p2Op2F expression vector for insect cells
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (561)..(566)
OTHER INFORMATION: HindIII site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (573)..(578)
OTHER INFORMATION: AvalI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (586)..(591)
OTHER INFORMATION: EcoRI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (593)..(598)
OTHER INFORMATION: BamHI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (625)..(630)
OTHER INFORMATION: ClaI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (629)..(634)
OTHER INFORMATION: ApaLI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (1156)..(1161)
OTHER INFORMATION: ApaLI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2128)..(2133)
OTHER INFORMATION: PstI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2204)..(2209)
OTHER INFORMATION: NcoI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2284)..(2289)
OTHER INFORMATION: AvalI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2294)..(2299)
OTHER INFORMATION: AvalI, SmaI, and XmaI site
FEATURE:
NAME/KEY: misc_recomb
LOCATION: (2551)..(2556)
OTHER INFORMATION: ApaLI site
US-10-939-107-60

Query Match 100.0%; Score 177; DB 24; Length 2773;
Best Local Similarity 100.0%; Pred. No. 2.6e-54;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCTTATCGTGACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGGACA 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
355 GTCTTATCGTGACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGA 414
Qy 61 TATCGGAACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGGACA 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
415 TATCGGAACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGGACA 474
Qy 121 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAA 177
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
475 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAA 531

RESULT 7

US-10-622-088-89
; Sequence 89, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth
; APPLICANT: Franke, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622,088
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/22437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 89
; LENGTH: 5038
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pIB/V5-His-DEST
US-10-622-088-89

Query Match 99.1%; Score 175.4; DB 21; Length 5038;
Best Local Similarity 99.4%; Pred. No. 1.1e-53;
Matches 176; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCTTATCGTGACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGA 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
351 GTCTTATCGTGACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGA 410
Qy 61 TATCGGAACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGGACA 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
411 TATCGGAACAGGAGCGCGCTTCATATCAGCGCGCGTTATCTCATGCGCGTGACCGGACA 470
Qy 121 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAA 177
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
471 CGAGCGCGCGTCCCGCTTATCGCGCTTATAAATACAGCGCGCAACGATCTGGTAAA 527

RESULT 8

US-10-622-088-127
; Sequence 127, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:

```
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth
; APPLICANT: Franke, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622,088
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/22437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 127
; LENGTH: 147
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombination region of pIB/V5-His-DEST
; NAME/KEY: misc_feature
; LOCATION: (141)..(148)
; OTHER INFORMATION: n may be any nucleotide
US-10-622-088-127

Query Match      23.2%; Score 41; DB 21; Length 147;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      137 CTTATCGCGCCTATAAATACAGCCGCGCAACGATCTGGTAAA 177
Db      1 CTTATCGCGCCTATAAATACAGCCGCGCAACGATCTGGTAAA 41

RESULT 9
US-10-622-088-149
; Sequence 149, Application US/10622088
; Publication No. US20040219516A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Robert P.
; APPLICANT: Welch, Peter J.
; APPLICANT: Harwood, Steven
; APPLICANT: Madden, Knut
; APPLICANT: Frimpong, Kenneth
; APPLICANT: Franke, Kenneth E.
; TITLE OF INVENTION: Viral Vectors Containing Recombination Sites
; FILE REFERENCE: 0942.5450007
; CURRENT APPLICATION NUMBER: US/10/622,088
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: PCT/US03/22437
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/396,335
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 60/398,617
; PRIOR FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 60/427,231
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/456,496
; PRIOR FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/474,940
; PRIOR FILING DATE: 2003-06-03
; NUMBER OF SEQ ID NOS: 165
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 149
; LENGTH: 325
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Recombination region of pIB/V5 His DEST
; NAME/KEY: misc_feature
; LOCATION: (141)..(142)
; OTHER INFORMATION: n is a, c, g, or t
; NAME/KEY: CDS
; LOCATION: (145)..(276)
US-10-622-088-149

Query Match      23.2%; Score 41; DB 21; Length 325;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      137 CTTATCGCGCCTATAAATACAGCCGCGCAACGATCTGGTAAA 177
Db      1 CTTATCGCGCCTATAAATACAGCCGCGCAACGATCTGGTAAA 41

RESULT 10
US-10-156-761-2538/c
; Sequence 2538, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2538
; LENGTH: 3138
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3138)
US-10-156-761-2538

Query Match      18.2%; Score 32.2; DB 16; Length 3138;
Best Local Similarity 61.2%; Pred. No. 0.3;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY      58 CTTATCGGAACAGGACGGCTCCATATACAGCCGCGCTTATCTCATCGCGGTGACCGG 117
Db      2575 CTTCCACCGAACCGGACGGCTCGCCCTTGATGATGAGGTGAGTTCTCCTGCACGACCGG 2516

QY      118 ACACGAGGCGCCGCTCCGCTTATC 142
Db      2515 CTTGAGGGCCTCGTCCAGGTCTC 2491

RESULT 11
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
```


Db 657 ACAACAGCTTCGGTGGAGGTCCCAACAGCGGGTCTAGTCTCTTGGCGGCTCCGGTGGC 716
QY 77 GCCTCCATATACGCCGCGGTTCATCTCATGCGCGTGACCGGACAGC 122
Db 717 GTTGCACATTGCCCATACAAATCTCAAGTCCAGAAAGTACAAG 762

RESULT 14

US-11-097-143-38495
; Sequence 38495, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38495
; LENGTH: 1342
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-38495

Query Match 17.4%; Score 30.8; DB 26; Length 1342;
Best Local Similarity 55.7%; Pred. No. 0.84;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 17 ACGCCAGCTTCCTGTGTTGCTAACCGCAGCGCGGAACTCCTTATCGGAACAGGACGC 76
Db 731 ACAACAGCTTCGGTGGAGGTCCCAACAGCGGGTCTAGTCTCTTGGCGGCTCCGGTGGC 790
QY 77 GCCTCCATATACGCCGCGGTTCATCTCATGCGCGTGACCGGACAGC 122
Db 791 GTTGCACATTGCCCATACAAATCTCAAGTCCAGAAAGTACAAG 836

RESULT 15

US-11-097-143-38494/c
; Sequence 38494, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191

; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38494
; LENGTH: 3448
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-38494
Query Match 17.4%; Score 30.8; DB 26; Length 3448;
Best Local Similarity 55.7%; Pred. No. 0.98;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 17 ACGCCAGCTTCCTGTGTTGCTAACCGCAGCGCGGAACTCCTTATCGGAACAGGACGC 76
Db 1667 ACAACAGCTTCGGTGGAGGTCCCAACAGCGGGTCTAGTCTCTTGGCGGCTCCGGTGGC 1608
QY 77 GCCTCCATATACGCCGCGGTTCATCTCATGCGCGTGACCGGACAGC 122
Db 1607 GTTGCACATTGCCCATACAAATCTCAAGTCCAGAAAGTACAAG 1562

Search completed: October 25, 2005, 06:13:38
Job time : 404.827 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:41:53 ; Search time 60.1934 Seconds
(without alignments)
4811.505 Million cell updates/sec

Title: US-09-896-888A-1_COPY_351_527

Perfect score: 177

Sequence: 1 gcttatcgtgacagcgc.....gcccgcaacgacatcgtgtaaa 177

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	30	16.9	1710	4	US-09-252-991A-12692
c 2	29.8	16.8	1311	4	US-09-902-540-2371
c 3	29.8	16.8	13579	4	US-09-902-540-1101
c 4	28.6	16.2	585	4	US-09-252-991A-12370
c 5	28.6	16.2	1443	4	US-09-252-991A-12097
c 6	28.6	16.2	1551	4	US-09-252-991A-12323
c 7	28.6	16.2	2085	4	US-09-252-991A-12237
c 8	28.6	16.2	2091	3	US-08-899-437-22
c 9	28.6	16.2	2091	3	US-08-126-121-22
c 10	28.6	16.2	2502	3	US-08-899-437-5
c 11	28.6	16.2	2502	3	US-09-126-121-5
c 12	28.6	16.2	25048	4	US-09-902-540-1239
c 13	28.4	16.0	57320	4	US-09-949-016-13983
c 14	28.2	15.9	283	4	US-09-513-998C-27806
c 15	28.2	15.9	2892	4	US-09-902-540-7217
c 16	28.2	15.9	3612	4	US-09-902-540-674
c 17	28.2	15.9	126200	4	US-09-949-016-11824
c 18	28.2	15.9	126200	4	US-09-949-016-13193
c 19	27.8	15.7	601	4	US-09-949-016-19425
c 20	27.8	15.7	601	4	US-09-949-016-49606
c 21	27.8	15.7	6208	4	US-09-902-540-7028
c 22	27.8	15.7	5588	4	US-09-902-540-621
c 23	27.6	15.6	534	4	US-09-489-039A-2740
c 24	27.6	15.6	798	4	US-09-489-039A-2763
c 25	27.4	15.5	37838	4	US-09-949-016-12134
c 26	27.2	15.4	549	4	US-09-902-540-6433
c 27	27.2	15.4	978	4	US-09-252-991A-14837

c 28	27.2	15.4	999	4	US-09-252-991A-14970	Sequence 14970, A
c 29	27.2	15.4	1377	4	US-09-902-540-2474	Sequence 2474, Ap
c 30	27.2	15.4	1524	4	US-09-489-039A-5551	Sequence 5551, Ap
c 31	27.2	15.4	4194	4	US-09-902-540-2099	Sequence 2099, Ap
c 32	27.2	15.4	4902	4	US-09-902-540-5635	Sequence 5635, Ap
c 33	27.2	15.4	8563	4	US-09-902-540-3318	Sequence 3318, Ap
c 34	27.2	15.4	9080	4	US-09-902-540-1363	Sequence 1363, Ap
c 35	27.2	15.4	12865	4	US-09-902-540-1048	Sequence 1048, Ap
c 36	27.2	15.4	15209	4	US-09-902-540-1110	Sequence 1110, Ap
c 37	27.2	15.4	15351	4	US-09-902-540-1154	Sequence 1154, Ap
c 38	27.2	15.4	49225	4	US-09-902-540-1269	Sequence 1269, Ap
c 39	27	15.3	1182	4	US-09-252-991A-200	Sequence 200, App
c 40	27	15.3	1230	4	US-10-105-319-1	Sequence 1, Appl
c 41	27	15.3	1698	4	US-09-252-991A-193	Sequence 193, App
c 42	27	15.3	1701	4	US-09-252-991A-219	Sequence 219, App
c 43	27	15.3	1938	4	US-09-252-991A-228	Sequence 228, App
c 44	27	15.3	20093	4	US-09-949-016-15207	Sequence 15207, A
c 45	27	15.3	77772	4	US-09-949-016-17417	Sequence 17417, A

ALIGNMENTS

RESULT 1

US-09-252-991A-12692/c
; Sequence 12692, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12692
; LENGTH: 1710
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12692

Query Match	16.9%	Score 30;	DB 4;	Length 1710;
Best Local Similarity	57.4%	Pred. No. 2;		
Matches	54;	Conservative	0;	Mismatches 40;
			Indels	0;
Gaps	0;			
QY	81	CCATATCAGCCGCGTTCATTCGCGTGACGCGACAGCGCCGTCGCCGTTA	140	
Db	414	CCAAATAGCCGCTCGATCTCTCGCGAAGTGACGGGCGACCTGCGCGCATTTGCCCC	355	
QY	141	TGCGCCCTTAATATACAGCCGCGCAACGATCTGGT	174	
Db	354	TGCGCCGCTCAACTCTCGATCCACCTGCGAGGT	321	

RESULT 2

US-09-902-540-2371
; Sequence 2371, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldmann, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wigand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10

```
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 2371
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-2371

Query Match      16.8%; Score 29.8; DB 4; Length 1311;
Best Local Similarity 55.2%; Pred. No. 2.1;
Matches 58; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 45 GCCGAGCGCAACTCTCTTATCGGAACAGGACGCGCTCCATATCAGCCGCGGTTATCTCA 104
Db 368 CCGCGAGTTCCTGCTGCTCTGGACGGAGACACCTCATCTTCTGACGGTGTCAACC 427

Qy 105 TGC CGCGTGACGGACAGGACGCGCGCTCCGCTTATCGCGCCTA 149
Db 428 GGAACGGGACATTGGACGCGCGCGGTACGCTCTTTACCCCA 472

RESULT 3
US-09-902-540-1101/c
; Sequence 1101, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1101
; LENGTH: 13579
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1101

Query Match      16.8%; Score 29.8; DB 4; Length 13579;
Best Local Similarity 55.2%; Pred. No. 4.3;
Matches 58; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 45 GCCGAGCGCAACTCTCTTATCGGAACAGGACGCGCTCCATATCAGCCGCGGTTATCTCA 104
Db 1063 CCGCGAGTTCCTGCTGCTCTGGACGGAGACACCTCCATCTTCTGACGGTGTCAACC 1004

Qy 105 TGC CGCGTGACGGACAGGACGCGCGCTCCGCTTATCGCGCCTA 149
Db 1003 GGAACGGGACATTGGACGCGCGCGGTACGCTCTTTACCCCA 959

RESULT 4
US-09-902-540-12370/c
; Sequence 12370, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12370
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12370

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12370

Query Match      16.2%; Score 28.6; DB 4; Length 585;
Best Local Similarity 53.0%; Pred. No. 4.3;
Matches 61; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 8 CGTGACAGGACGCGCCTTCTCTGTTGTAAACCGGACGCGGACGCAACTCTTATCGGA 67
Db 534 CGTGCCTGCTGGCAGGGGCTGTGTCGACGCGCATGCCGCTTCGATACCGTGGTCA 475

Qy 68 ACAGGACGCGCCTCATATCAGCGCGCGTTATCTCATGCGCGTACCGGACACG 122
Db 474 GCTGGACGCGCGCAGATCAAGCGCGAGGTGAGCTGGGGCACGTCGCGGAAATG 420

RESULT 5
US-09-252-991A-12097
; Sequence 12097, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12097
; LENGTH: 1443
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12097

Query Match      16.2%; Score 28.6; DB 4; Length 1443;
Best Local Similarity 53.0%; Pred. No. 5.7;
Matches 61; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 8 CGTGACAGGACGCGCCTTCTCTGTTGTAAACCGGACGCGGACGCAACTCTTATCGGA 67
Db 810 CGTGCCTGCTGGCAGGGGCTGTGTCGACGCGCATGCCGCTTCGATACCGTGGTCA 869

Qy 68 ACAGGACGCGCCTCATATCAGCGCGCGTTATCTCATGCGCGTACCGGACACG 122
Db 870 GCTGGACGCGCGCAGATCAAGCGCGAGGTGAGCTGGGGCACGTCGCGGAAATG 924

RESULT 6
US-09-252-991A-12323/c
; Sequence 12323, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12323
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12323
```

```

Query Match      16.2%; Score 28.6; DB 4; Length 2085;
Best Local Similarity 53.0%; Pred.No. 6.3;
Matches      61; Conservative      0; Mismatches 54; Indels      0; Gaps      0;

QY      8  CGTCACAGGACGCCAGCTTCCTGTGTGCTAAACGGACGCGGACGGAACCTCCTATCGGA 67
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      802  CGTCGCTCTGTGCAGGGCTGGTGTGCGAGCGCGATGCCCGCTTCGTATCCGTGGTGTGCA 861

QY      68  ACAGGACGCGCTCATATCAGCGCGCGCTTATCTCATGGCGGTGACCGGACACG 122
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      862  GCTGGACGCGCGCAGATCAAGCGCAGGTCACTGGGGCACGTCGCCGGAATG 916

RESULT 8
US-08-899-437-22
: Sequence 22, Application US/08899437
: Patent No. 612415
: GENERAL INFORMATION:
: APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
: TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
: TITLE OF INVENTION: Ligands and Uses Thereof
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 1 DNA Way
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WinPatIn (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/899,437
: FILING DATE: 24-Jul-1997

```


Db 1657 CAGGTGACGGCTTCGGGGTTGCTCGCTCCAGCCCAACGCCGGAAGCCAGGGGGAGTAC 1716

Qy 73 ACGGCTTCATATACAGCCGCGTTATCTCATGCGGTGACCGG 117

Db 1717 GCGGCTGCTCGTCATCCGCGGTACCAACAGAGCCGCGCCAG 1761

Search completed: October 24, 2005, 21:57:49
Job time : 63.1934 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:40:53 ; Search time 302.908 Seconds
(without alignments)
7998.346 Million cell updates/sec

Title: US-09-896-888A-10

Perfect score: 50

Sequence: 1 acttaagctttatagcatga.....tccagtcgggaacctgtcg 50

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	50	100.0	50	6	BD070868	Insect ex
2	36.8	73.6	99591	2	AC013392	Homo sapi
3	36.8	73.6	118229	2	AC013315	Homo sapi
4	36.8	73.6	211967	2	AC013368	Homo sapi
5	36.8	73.6	260636	2	AC015852	Homo sapi
6	36.2	72.4	73282	2	AC010800	Homo sapi
7	36.2	72.4	217141	2	AC073998	Homo sapi
8	35.8	71.6	604	11	G39061	BD269301 33 human
9	35.8	71.6	826	11	G40613	G40613 27956 Zebra
10	35.8	71.6	979	11	G40162	G40162 220177 Zebra
11	35.8	71.6	76295	2	AC023542	Homo sapi
12	35.6	71.2	586	6	CQ406071	Sequence
13	35.6	71.2	75002	2	AC023384	Homo sapi
14	35.6	71.2	95127	2	AC016798	Homo sapi
15	35.6	71.2	115468	2	AC011820	Homo sapi
16	35.6	71.2	170427	2	AC006095	Homo sapi
17	35.6	71.2	170427	2	AC006095	Homo sapi
18	35.2	70.4	114	6	AR092751	Sequence
19	35.2	70.4	114	6	AR359133	Sequence

20	35.2	70.4	173	6	CQ504404	Sequence
21	35.2	70.4	179	6	CQ513399	Sequence
22	35.2	70.4	186	6	E00523	Double-str
23	35.2	70.4	196	6	AX100258	Sequence
24	35.2	70.4	201	6	AR038286	Sequence
25	35.2	70.4	201	6	AR075283	Sequence
26	35.2	70.4	201	6	AR152695	Sequence
27	35.2	70.4	201	6	I58575	Sequence 41
28	35.2	70.4	201	6	I61307	Sequence 11
29	35.2	70.4	201	6	AR238535	Sequence
30	35.2	70.4	201	6	AX323349	Sequence
31	35.2	70.4	203	6	CQ504346	Sequence
32	35.2	70.4	205	6	AX361570	Sequence
33	35.2	70.4	207	6	AI5078	Lac promote
34	35.2	70.4	213	6	BD175852	method
35	35.2	70.4	213	6	AX404725	Sequence
36	35.2	70.4	221	6	I92483	Sequence 1
37	35.2	70.4	226	6	I03664	Sequence 2
38	35.2	70.4	241	11	G02286	human STS S
39	35.2	70.4	244	6	CQ513678	Sequence
40	35.2	70.4	253	6	CQ506464	Sequence
41	35.2	70.4	255	8	CAR535749	Cicer ari
42	35.2	70.4	258	12	SYNML13MP9C	Cloning vec
43	35.2	70.4	286	7	M13LACMUT	Bacterioph
44	35.2	70.4	304	6	AR527545	Sequence
45	35.2	70.4	313	12	SYNECONUTX	Cloning vec

ALIGNMENTS

RESULT 1
BD070868
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BD070868
Insect expression vectors.
BD070868
BD070868.1 GI:22616471
JP 2001516225-A/13.
unidentified
unidentified
unclassified.
1 (bases 1 to 50)
Grigliatti,T.A., Theilmann,D.A., Pfeifer,T.A. and Hegedus,D.D.
Insect expression vectors
Patent: JP 2001516225-A 13 25-SEP-2001;
THE UNIVERSITY OF BRITISH COLUMBIA
OS Unidentified
PN JP 2001516225-A/13
PD 25-SEP-2001
PF 26-MAR-1998 JP 1998541010
PR 27-MAR-1997 US 60/049946,28-JAN-1998 CA 2221819 PI
THOMAS A GRIGLIATTI,DAVE A THEILMANN,THOMAS
A PFEIFER,DWAYNE D
PI HEGEDUS
PC C12N15/85,C12N5/06,C12N15/69//C12N9/22
CC Insect expression vectors
FH Key Location/Qualifiers
FT source 1..50
FT /organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..50
/organism='unidentified'
/mol_type='genomic DNA'
/db_xref='taxon:32644'

Query Match 100.0%; Score 50; DB 6; Length 50;
Best Local Similarity 100.0%; Pred.No. 2e-08;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ACTTAAGCTTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAACCTGTGCG 50
|||||
Db 1 ACTTAAGCTTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAACCTGTGCG 50
|||||

RESULT 2

AC013392

LOCUS

AC013392 Homo sapiens chromosome 2 clone RP11-429N24 map 2, LOW-PASS
SEQUENCE SAMPLING.

ACCESSION

AC013392.3 GI:9123920

VERSION

HTG; HTGS_PHASE0.

SOURCE

Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens chromosome 2, clone RP11-429N24

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 99591)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6425709.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIPR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I3805
Center clone name: 429_N_24

* NOTE: This record contains 111 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 827: contig of 827 bp in length
* 828 927: gap of 100 bp
* 928 1726: contig of 799 bp in length
* 1727 1826: gap of 100 bp
* 1827 2635: contig of 809 bp in length
* 2636 2735: gap of 100 bp
* 2736 3547: contig of 812 bp in length
* 3548 3647: gap of 100 bp
* 3648 4424: contig of 777 bp in length
* 4425 5303: contig of 779 bp in length
* 5304 5403: gap of 100 bp
* 5404 6171: contig of 768 bp in length
* 6172 6271: gap of 100 bp
* 6272 7071: contig of 800 bp in length
* 7072 7171: gap of 100 bp
* 7172 7979: contig of 808 bp in length
* 7980 8079: gap of 100 bp
* 8080 8881: contig of 802 bp in length
* 8882 8981: gap of 100 bp
* 8982 9732: contig of 751 bp in length
* 9733 9832: gap of 100 bp
* 9833 10559: contig of 727 bp in length
* 10560 11463: contig of 804 bp in length
* 11464 11563: gap of 100 bp
* 11564 12340: contig of 777 bp in length
* 12341 12440: gap of 100 bp
* 12442 13237: contig of 797 bp in length
* 13238 13337: gap of 100 bp
* 13339 14134: contig of 797 bp in length
* 14140 14234: gap of 100 bp
* 14235 15002: contig of 768 bp in length
* 15003 15102: gap of 100 bp
* 15103 15872: contig of 770 bp in length
* 15873 15972: gap of 100 bp
* 15973 16731: contig of 759 bp in length
* 16732 17605: contig of 774 bp in length
* 17606 17705: gap of 100 bp
* 17706 18499: contig of 794 bp in length
* 18500 18599: gap of 100 bp
* 18600 19401: contig of 802 bp in length
* 19402 19501: gap of 100 bp
* 19502 20325: contig of 824 bp in length
* 20326 21189: contig of 764 bp in length
* 21190 22048: contig of 759 bp in length
* 22049 22148: gap of 100 bp
* 22149 22907: contig of 759 bp in length
* 22908 23007: gap of 100 bp
* 23008 23776: contig of 768 bp in length
* 23776 24684: contig of 809 bp in length
* 24685 24784: gap of 100 bp
* 24785 25560: contig of 776 bp in length
* 25561 25660: gap of 100 bp
* 25661 26447: contig of 786 bp in length
* 26447 26547: contig of 776 bp in length
* 26547 27323: gap of 100 bp
* 27323 28179: contig of 757 bp in length
* 28180 28279: gap of 100 bp
* 28280 29044: contig of 765 bp in length
* 29045 29144: gap of 100 bp
* 29145 29942: contig of 798 bp in length
* 29943 30042: gap of 100 bp
* 30043 30811: contig of 769 bp in length
* 30812 30911: gap of 100 bp
* 30912 31663: contig of 752 bp in length
* 31664 31763: gap of 100 bp
* 31764 32540: contig of 777 bp in length
* 32541 32640: gap of 100 bp
* 32641 33471: contig of 831 bp in length
* 33472 33571: gap of 100 bp
* 33572 34358: contig of 787 bp in length
* 34359 34458: gap of 100 bp
* 34459 35239: contig of 781 bp in length
* 35240 35339: gap of 100 bp
* 35340 36111: contig of 772 bp in length
* 36112 36211: gap of 100 bp
* 36212 36966: contig of 755 bp in length
* 36967 37066: gap of 100 bp
* 37067 37813: contig of 747 bp in length
* 37814 37913: gap of 100 bp

```

* 37914 38694: contig of 781 bp in length
* 38695 38794: gap of 100 bp
* 38795 39559: contig of 765 bp in length
* 39560 39659: gap of 100 bp
* 39660 40413: contig of 754 bp in length
* 40414 40513: gap of 100 bp
* 40514 41335: contig of 822 bp in length
* 41336 42174: gap of 100 bp
* 41436 42174: contig of 739 bp in length
* 42175 42275: gap of 100 bp
* 42275 43080: contig of 806 bp in length
* 43081 43181: gap of 100 bp
* 43181 43371: contig of 791 bp in length
* 43372 44071: gap of 100 bp
* 44072 44876: contig of 805 bp in length
* 44877 44976: gap of 100 bp
* 44977 45791: contig of 815 bp in length
* 45792 45892: gap of 100 bp
* 45892 46720: contig of 829 bp in length
* 46721 46820: gap of 100 bp
* 46821 47558: contig of 738 bp in length
* 47559 48475: gap of 100 bp
* 48475 48575: contig of 817 bp in length
* 48576 49384: gap of 100 bp
* 49385 49484: gap of 100 bp
* 49485 50346: contig of 862 bp in length
* 50347 50446: gap of 100 bp
* 50447 51295: contig of 849 bp in length
* 51296 51395: gap of 100 bp
* 51396 52202: contig of 807 bp in length
* 52203 53120: gap of 100 bp
* 53120 53220: contig of 818 bp in length
* 53221 53220: gap of 100 bp
* 53221 54050: contig of 830 bp in length
* 54051 54150: gap of 100 bp
* 54151 54967: contig of 817 bp in length
* 54968 55067: gap of 100 bp
* 55068 55868: contig of 801 bp in length
* 55869 55968: gap of 100 bp
* 55969 56805: contig of 837 bp in length
* 56806 56905: gap of 100 bp
* 56906 57747: contig of 842 bp in length
* 57748 57847: gap of 100 bp
* 57848 58650: contig of 803 bp in length
* 58651 58750: gap of 100 bp
* 58751 59553: contig of 803 bp in length
* 59554 59654: gap of 100 bp
* 59654 60386: contig of 733 bp in length
* 60387 60486: gap of 100 bp
* 60487 61311: contig of 825 bp in length
* 61312 61411: gap of 100 bp
* 61412 62225: contig of 814 bp in length
* 62226 62325: gap of 100 bp
* 62326 63128: contig of 803 bp in length
* 63129 63228: gap of 100 bp
* 63229 64055: contig of 827 bp in length
* 64056 64155: gap of 100 bp

```

```

Query Match 73.6%; Score 36.8; DB 2; Length 99591;
Best Local Similarity 85.4%; Pred. No. 0.0031;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TTAACCTTATACGATGACTGCCCTTTCAGTCGGGAACCTCTCG 50
DB 3876 TTAATGCGTGGCTGACTGCCCTTTCAGTCGGGAACCTCTCG 3923

```

```

RESULT 3
AC013315
LOCUS AC013315 118229 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 22 clone RP11-348B6 map 22, LOW-PASS
SEQUENCE SAMPLING.

```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

AC013315
AC013315.3 GI:9123904
HTG: HTGS_PHASE0.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 118229)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 22, clone RP11-348B6
Unpublished
2 (bases 1 to 118229)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collins,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melidri,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 118229)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campolano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meidrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced GI:6425714.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3764
Center clone name: 348_B_6

TITLE JOURNAL COMMENT

* NOTE: This record contains 124 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 765: contig of 765 bp in length
* 766 865: gap of 100 bp
* 866 1676: contig of 811 bp in length
* 1677 1776: gap of 100 bp
* 1777 2577: contig of 801 bp in length
* 2578 2677: gap of 100 bp
* 2678 3519: contig of 842 bp in length
* 3520 3619: gap of 100 bp
* 3620 4441: contig of 822 bp in length
* 4442 4541: gap of 100 bp
* 4542 5305: contig of 764 bp in length
* 5306 5405: gap of 100 bp
* 5406 6261: contig of 856 bp in length
* 6262 6361: gap of 100 bp
* 6362 7140: contig of 779 bp in length
* 7141 7240: gap of 100 bp
* 7241 8021: contig of 781 bp in length
* 8022 8121: gap of 100 bp
* 8122 8925: contig of 804 bp in length
* 8926 9025: gap of 100 bp
* 9026 9840: contig of 815 bp in length
* 9841 9940: gap of 100 bp
* 9941 10717: contig of 777 bp in length
* 10718 10817: gap of 100 bp
* 10818 11617: contig of 800 bp in length
* 11618 11717: gap of 100 bp
* 11718 12526: contig of 809 bp in length
* 12527 12626: gap of 100 bp
* 12627 13506: contig of 880 bp in length
* 13507 13606: gap of 100 bp
* 13607 14393: contig of 787 bp in length
* 14394 14493: gap of 100 bp
* 14494 15276: contig of 783 bp in length
* 15277 13776: gap of 100 bp
* 15278 16162: contig of 786 bp in length
* 16163 16262: gap of 100 bp
* 16263 17045: contig of 783 bp in length
* 17046 17145: gap of 100 bp
* 17146 17966: contig of 821 bp in length
* 17967 18066: gap of 100 bp
* 18067 18866: contig of 800 bp in length
* 18867 18966: gap of 100 bp
* 18967 19853: contig of 887 bp in length
* 19854 19954: gap of 100 bp
* 19954 20765: contig of 812 bp in length
* 20766 20865: gap of 100 bp
* 20866 21643: contig of 778 bp in length
* 21644 21743: gap of 100 bp
* 21744 22580: contig of 837 bp in length
* 22581 22680: gap of 100 bp
* 22681 23500: contig of 820 bp in length
* 23501 23600: gap of 100 bp
* 23601 24406: contig of 806 bp in length
* 24407 24506: gap of 100 bp
* 24507 25329: contig of 823 bp in length
* 25330 25429: gap of 100 bp
* 25430 26242: contig of 813 bp in length
* 26243 26342: gap of 100 bp
* 26343 27154: contig of 812 bp in length
* 27155 27254: gap of 100 bp
* 27255 28036: contig of 782 bp in length
* 28037 28136: gap of 100 bp
* 28137 28937: contig of 801 bp in length
* 28938 29037: gap of 100 bp
* 29038 29798: contig of 761 bp in length
* 29799 29898: gap of 100 bp
* 29899 30725: contig of 827 bp in length
* 30726 30825: gap of 100 bp

* 30826 31709: contig of 884 bp in length
* 31710 31809: gap of 100 bp
* 31810 32621: contig of 812 bp in length
* 32622 32721: gap of 100 bp
* 32722 33508: contig of 787 bp in length
* 33509 33608: gap of 100 bp
* 33609 34400: contig of 792 bp in length
* 34401 34500: gap of 100 bp
* 34501 35360: contig of 860 bp in length
* 35361 35460: gap of 100 bp
* 35461 36284: contig of 824 bp in length
* 36285 36384: gap of 100 bp
* 36385 37168: contig of 784 bp in length
* 37169 37268: gap of 100 bp
* 37269 38067: contig of 799 bp in length
* 37270 38167: gap of 100 bp
* 38068 38961: contig of 794 bp in length
* 38962 39061: gap of 100 bp
* 39062 39876: contig of 815 bp in length
* 39877 39976: gap of 100 bp
* 39977 40789: contig of 813 bp in length
* 40790 40889: gap of 100 bp
* 40890 41741: contig of 852 bp in length
* 41742 41841: gap of 100 bp
* 41842 42695: contig of 854 bp in length
* 42696 42795: gap of 100 bp
* 42796 43665: contig of 870 bp in length
* 43666 43765: gap of 100 bp
* 43766 44665: contig of 900 bp in length
* 44666 44765: gap of 100 bp
* 44766 45661: contig of 896 bp in length
* 45662 45761: gap of 100 bp
* 45762 46649: contig of 888 bp in length
* 46650 46749: gap of 100 bp
* 46750 47621: contig of 872 bp in length
* 47622 47721: gap of 100 bp
* 47722 48586: contig of 865 bp in length
* 48587 48686: gap of 100 bp
* 48687 49538: contig of 852 bp in length
* 49539 49639: gap of 100 bp
* 49640 50483: contig of 845 bp in length
* 50484 51438: contig of 855 bp in length
* 51439 51538: gap of 100 bp
* 51539 52399: contig of 861 bp in length
* 52400 52499: gap of 100 bp
* 52500 53388: contig of 889 bp in length
* 53389 53488: gap of 100 bp
* 53489 54367: contig of 879 bp in length
* 54368 54467: gap of 100 bp
* 54468 55344: contig of 877 bp in length

Query Match 73.6%; Score 36.8; DB 2; Length 118229;
Best Local Similarity 85.4%; Pred. No. 0.0031;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TTATGCTTATAGCATCTCCCGCTTCCAGTCGGGAACCTGTCG 50
|||||
Db 9250 TTAATTCGTGCGCTGACTCCCGCTTCCAGTCGGGAACCTGTCG 9297

RESULT 4

AC013368
LOCUS AC013368 211967 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-11016, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC013368
VERSION AC013368.6 GI:9123839
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 211967)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-11016
Unpublished
2 (bases 1 to 211967)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barns,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castelli,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,C., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:8099785.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3239
Center clone name: 11_Q_16

TITLE
JOURNAL
COMMENT

* NOTE: This record contains 254 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 637: contig of 637 bp in length
* 638 737: gap of 100 bp
* 738 1380: contig of 643 bp in length
* 1381 1480: gap of 100 bp
* 1481 2102: contig of 622 bp in length
* 2103 2202: gap of 100 bp
* 2203 2824: contig of 622 bp in length
* 2825 2924: gap of 100 bp
* 2925 3591: contig of 667 bp in length
* 3592 3691: gap of 100 bp
* 3692 4364: contig of 673 bp in length
* 4365 4464: gap of 100 bp
* 4465 5118: contig of 654 bp in length
* 5119 5218: gap of 100 bp
* 5219 5879: contig of 661 bp in length
* 5880 5979: gap of 100 bp
* 5980 6655: contig of 676 bp in length
* 6656 7382: gap of 100 bp
* 7383 7482: contig of 627 bp in length
* 7483 8135: contig of 653 bp in length
* 8136 8235: gap of 100 bp
* 8236 8900: contig of 665 bp in length
* 8901 9000: gap of 100 bp
* 9001 9652: contig of 652 bp in length
* 9653 9752: gap of 100 bp
* 9753 10378: contig of 626 bp in length
* 10379 10478: gap of 100 bp
* 10479 11122: contig of 644 bp in length
* 11123 11222: gap of 100 bp
* 11223 11880: contig of 658 bp in length
* 11881 11980: gap of 100 bp
* 11981 12540: contig of 660 bp in length
* 12541 12740: gap of 100 bp
* 12741 13400: contig of 660 bp in length
* 13401 13500: gap of 100 bp
* 13501 14140: contig of 640 bp in length
* 14141 14240: gap of 100 bp
* 14241 14876: contig of 636 bp in length
* 14877 14976: gap of 100 bp
* 14977 15635: contig of 659 bp in length
* 15636 15735: gap of 100 bp
* 15736 16390: contig of 655 bp in length
* 16391 16490: gap of 100 bp
* 16491 17143: contig of 653 bp in length
* 17144 17243: gap of 100 bp
* 17244 17886: contig of 643 bp in length
* 17887 17986: gap of 100 bp
* 17987 18628: contig of 642 bp in length
* 18629 18728: gap of 100 bp
* 18729 19383: contig of 655 bp in length
* 19384 19483: gap of 100 bp
* 19484 20165: contig of 682 bp in length
* 20166 20265: gap of 100 bp
* 20266 20906: contig of 641 bp in length
* 20907 21006: gap of 100 bp
* 21007 21580: contig of 674 bp in length
* 21581 21780: gap of 100 bp
* 21781 22410: contig of 630 bp in length
* 22411 22510: gap of 100 bp
* 22511 23164: contig of 654 bp in length
* 23165 23264: gap of 100 bp
* 23265 23927: contig of 663 bp in length
* 23928 24027: gap of 100 bp
* 24028 24685: contig of 658 bp in length
* 24686 24785: gap of 100 bp
* 24786 25427: contig of 642 bp in length
* 25428 25527: gap of 100 bp
* 25528 26193: contig of 666 bp in length
* 26194 28293: gap of 100 bp
* 28294 28952: contig of 659 bp in length
* 28953 27052: gap of 100 bp
* 27053 27723: contig of 671 bp in length
* 27724 27823: gap of 100 bp
* 27824 28468: contig of 645 bp in length
* 28469 28568: gap of 100 bp
* 28569 29228: contig of 660 bp in length
* 29229 29328: gap of 100 bp
* 29329 29976: contig of 648 bp in length
* 29977 30076: gap of 100 bp
* 30077 30732: contig of 656 bp in length
* 30733 30832: gap of 100 bp
* 30833 31471: contig of 639 bp in length
* 31472 31571: gap of 100 bp
* 31572 32210: contig of 639 bp in length
* 32211 32310: gap of 100 bp
* 32311 32950: contig of 640 bp in length
* 32951 33050: gap of 100 bp
* 33051 33706: contig of 656 bp in length
* 33707 33806: gap of 100 bp
* 33807 34434: contig of 628 bp in length
* 34435 34534: gap of 100 bp
* 34535 35188: contig of 654 bp in length
* 35189 35288: gap of 100 bp
* 35289 35952: contig of 664 bp in length
* 35953 36052: gap of 100 bp
* 36053 36683: contig of 631 bp in length
* 36684 36783: gap of 100 bp
* 36784 37452: contig of 669 bp in length
* 37453 37552: gap of 100 bp
* 37553 38208: contig of 656 bp in length

```

* 38209 38308: gap of 100 bp
* 38309 38967: contig of 659 bp in length
* 38968 39067: gap of 100 bp
* 39068 39722: contig of 655 bp in length
* 39723 39822: gap of 100 bp
* 39823 40412: contig of 590 bp in length
* 40413 40512: gap of 100 bp
* 40513 41155: contig of 643 bp in length
* 41156 41255: gap of 100 bp
* 41256 41925: contig of 670 bp in length
* 41926 42025: gap of 100 bp
* 42026 42695: contig of 670 bp in length
* 42696 42795: gap of 100 bp
* 42796 43442: contig of 647 bp in length
* 43443 43542: gap of 100 bp
* 43543 44165: contig of 623 bp in length
* 44166 44265: gap of 100 bp
* 44266 44933: contig of 668 bp in length
* 44934 45033: gap of 100 bp
* 45034 45670: contig of 637 bp in length
* 45671 45770: gap of 100 bp
* 45771 46415: contig of 645 bp in length
* 46416 46515: gap of 100 bp
* 46516 47172: contig of 657 bp in length
* 47173 47272: gap of 100 bp
* 47273 47944: contig of 672 bp in length
* 47945 48044: gap of 100 bp
* 48045 48682: contig of 638 bp in length
* 48683 48782: gap of 100 bp
* 48783 49424: contig of 642 bp in length
* 49425 49524: gap of 100 bp
* 49525 50188: contig of 664 bp in length
* 50189 50288: gap of 100 bp
* 50289 50958: contig of 670 bp in length
* 50959 51058: gap of 100 bp
* 51059 51720: contig of 662 bp in length
* 51721 51820: gap of 100 bp
* 51821 52470: contig of 650 bp in length
* 52471 52570: gap of 100 bp
* 52571 53112: contig of 742 bp in length
* 53113 53412: gap of 100 bp
* 53413 54150: contig of 738 bp in length
* 54151 54250: gap of 100 bp
* 54251 54986: contig of 736 bp in length

Query Match 73.6%; Score 35.8; DB 2; Length 211967;
Best Local Similarity 85.4%; Pred. No. 0.0032;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Qy 3 TTAAGCTTATAGCGATGACTGCCGCTTTCAGTCGGAAACCTGTCG 50
|||||
Db 130366 TTAAGTCGTTGCCGCTCACTACCGCGCTTTCAGTCGGAAACCTGTCG 130413
|||||

```

```

RESULT 5
AC015852
LOCUS AC015852 260636 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 18 clone CTB-155C15 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC015852
VERSION AC015852.2 Gi:9144626
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 260636)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone CTB-155C15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 260636)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,

```

```

Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczký,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 260636)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burckett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczký,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeeters,R.,
Meldrim,J., Meneus,L., Minova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6446816.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L518
Center clone name: 155_C_15
-----
* NOTE: This record contains 270 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 903: contig of 903 bp in length
* 904 1003: gap of 100 bp
* 1004 1879: contig of 876 bp in length
* 1880 1979: gap of 100 bp
* 1980 2862: contig of 883 bp in length
* 2863 2962: gap of 100 bp
* 2963 3845: contig of 887 bp in length
* 3850 3949: gap of 100 bp
* 3950 4832: contig of 883 bp in length

```

```

TITLE
JOURNAL
COMMENT

```

```
* 4833 4932: gap of 100 bp
* 4933 5810: contig of 878 bp in length
* 5811 5910: gap of 100 bp
* 5911 6799: contig of 889 bp in length
* 6800 6899: gap of 100 bp
* 6900 7777: contig of 878 bp in length
* 7778 7877: gap of 100 bp
* 7878 8449: contig of 872 bp in length
* 8450 8750: gap of 100 bp
* 8751 9266: contig of 877 bp in length
* 9267 9826: gap of 100 bp
* 9827 10715: contig of 883 bp in length
* 10716 10815: gap of 100 bp
* 10816 11687: contig of 872 bp in length
* 11688 11787: gap of 100 bp
* 11788 12664: contig of 877 bp in length
* 12665 12764: gap of 100 bp
* 12765 13644: contig of 880 bp in length
* 13645 13744: gap of 100 bp
* 13745 14626: contig of 882 bp in length
* 14627 14726: gap of 100 bp
* 14727 15612: contig of 886 bp in length
* 15613 15712: gap of 100 bp
* 15713 16600: contig of 888 bp in length
* 16601 16700: gap of 100 bp
* 16701 17572: contig of 872 bp in length
* 17573 17672: gap of 100 bp
* 17673 18554: contig of 882 bp in length
* 18555 18654: gap of 100 bp
* 18655 19333: contig of 879 bp in length
* 19334 19633: gap of 100 bp
* 19634 20498: contig of 865 bp in length
* 20499 20598: gap of 100 bp
* 20599 21473: contig of 875 bp in length
* 21474 21573: gap of 100 bp
* 21574 22462: contig of 889 bp in length
* 22463 22562: gap of 100 bp
* 22563 23439: contig of 877 bp in length
* 23440 23539: gap of 100 bp
* 23540 24404: contig of 865 bp in length
* 24405 24504: gap of 100 bp
* 24505 25380: contig of 876 bp in length
* 25381 25480: gap of 100 bp
* 25481 26372: contig of 892 bp in length
* 26373 26472: gap of 100 bp
* 26473 27361: contig of 889 bp in length
* 27362 27461: gap of 100 bp
* 27462 28340: contig of 879 bp in length
* 28341 28440: gap of 100 bp
* 28441 29317: contig of 877 bp in length
* 29318 29417: gap of 100 bp
* 29418 30303: contig of 886 bp in length
* 30304 30403: gap of 100 bp
* 30404 31284: contig of 881 bp in length
* 31285 31384: gap of 100 bp
* 31385 32255: contig of 871 bp in length
* 32256 32355: gap of 100 bp
* 32356 33235: contig of 880 bp in length
* 33236 33335: gap of 100 bp
* 33336 34222: contig of 887 bp in length
* 34223 34322: gap of 100 bp
* 34323 35197: contig of 875 bp in length
* 35198 35297: gap of 100 bp
* 35298 36164: contig of 867 bp in length
* 36165 36264: gap of 100 bp
* 36265 37144: contig of 880 bp in length
* 37145 37244: gap of 100 bp
* 37245 38114: contig of 870 bp in length
* 38115 38214: gap of 100 bp
* 38215 39104: contig of 890 bp in length
* 39105 39204: gap of 100 bp
* 39205 40089: contig of 885 bp in length
* 40090 40189: gap of 100 bp
```

```
* 40190 41061: contig of 872 bp in length
* 41062 41161: gap of 100 bp
* 41162 42037: contig of 876 bp in length
* 42038 42137: gap of 100 bp
* 42138 43012: contig of 875 bp in length
* 43013 43112: gap of 100 bp
* 43113 43980: contig of 868 bp in length
* 43981 44080: gap of 100 bp
* 44081 44959: contig of 879 bp in length
* 44960 45059: gap of 100 bp
* 45060 45943: contig of 884 bp in length
* 45944 46043: gap of 100 bp
* 46044 46917: contig of 874 bp in length
* 46918 47017: gap of 100 bp
* 47018 47894: contig of 877 bp in length
* 47895 47994: gap of 100 bp
* 47995 48870: contig of 876 bp in length
* 48871 48970: gap of 100 bp
* 48971 49844: contig of 874 bp in length
* 49845 49944: gap of 100 bp
* 49945 50826: contig of 882 bp in length
* 50827 50926: gap of 100 bp
* 50927 51806: contig of 880 bp in length
* 51807 51906: gap of 100 bp
* 51907 52777: contig of 871 bp in length
* 52778 52877: gap of 100 bp
* 52878 53760: contig of 883 bp in length
* 53761 53860: gap of 100 bp
* 53861 54727: contig of 867 bp in length
* 54728 54827: gap of 100 bp
* 54828 55699: contig of 872 bp in length
* 55700 55799: gap of 100 bp
* 55800 56689: contig of 890 bp in length
* 56690 56789: gap of 100 bp
* 56790 57663: contig of 874 bp in length
* 57664 57763: gap of 100 bp
* 57764 58633: contig of 870 bp in length

Query Match 73.6%; Score 36.8; DB 2; Length 260636;
Best Local Similarity 85.4%; Pred. No. 0.0032;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TTAAGCTTATAGCGATGACTCCCGCTTTCAGTTCGGAAACCTGTCG 50
    |||||
Db 128125 TTAATTGTGTGCGCTCACTCCCGCTTTCAGTTCGGAAACCTGTCG 128172

RESULT 6
AC010800
LOCUS Homo sapiens clone 1_E_17, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC010800
ACCESSION AC010800
VERSION AC010800.1 GI:5919335
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 73282)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone 1_E_17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 73282)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArallano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,E., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
```

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 859: contig of 859 bp in length
* gap of unknown length
* 860 1695: contig of 836 bp in length
* gap of unknown length
* 1696 2546: contig of 851 bp in length
* gap of unknown length
* 2547 3398: contig of 852 bp in length
* gap of unknown length
* 3399 4263: contig of 865 bp in length
* gap of unknown length
* 4264 5163: contig of 900 bp in length
* gap of unknown length
* 5164 6030: contig of 867 bp in length
* gap of unknown length
* 6031 6901: contig of 871 bp in length
* gap of unknown length
* 6902 7717: contig of 816 bp in length
* gap of unknown length
* 7718 8565: contig of 848 bp in length
* gap of unknown length
* 8566 9412: contig of 847 bp in length
* gap of unknown length
* 9413 10266: contig of 854 bp in length
* gap of unknown length
* 10267 11243: contig of 977 bp in length
* gap of unknown length
* 11244 12098: contig of 855 bp in length
* gap of unknown length
* 12099 12940: contig of 842 bp in length
* gap of unknown length
* 12941 13799: contig of 859 bp in length
* gap of unknown length
* 13800 14670: contig of 871 bp in length
* gap of unknown length
* 14671 15489: contig of 819 bp in length
* gap of unknown length
* 15490 16379: contig of 890 bp in length
* gap of unknown length
* 16380 17216: contig of 837 bp in length
* gap of unknown length
* 17217 18204: contig of 988 bp in length
* gap of unknown length
* 18205 19021: contig of 817 bp in length
* gap of unknown length
* 19022 19877: contig of 856 bp in length
* gap of unknown length
* 19878 20718: contig of 841 bp in length
* gap of unknown length
* 20719 21555: contig of 837 bp in length
* gap of unknown length
* 21556 22372: contig of 817 bp in length
* gap of unknown length
* 22373 23227: contig of 855 bp in length
* gap of unknown length
* 23228 24050: contig of 823 bp in length
* gap of unknown length
* 24051 24905: contig of 855 bp in length
* gap of unknown length
* 24906 25754: contig of 849 bp in length
* gap of unknown length
* 25755 26583: contig of 829 bp in length
* gap of unknown length
* 26584 27409: contig of 826 bp in length
* gap of unknown length
* 27410 28215: contig of 806 bp in length
* gap of unknown length
* 28216 29040: contig of 825 bp in length
* gap of unknown length
* 29041 29893: contig of 853 bp in length
* gap of unknown length
* 29894 30736: contig of 843 bp in length
* gap of unknown length
* 30737 31575: contig of 839 bp in length
* gap of unknown length
* 31576 32422: contig of 847 bp in length
* gap of unknown length
* 32423 33293: contig of 871 bp in length
* gap of unknown length
* 33294 34134: contig of 841 bp in length
* gap of unknown length
* 34135 34971: contig of 837 bp in length
* gap of unknown length
* 34972 35789: contig of 818 bp in length
* gap of unknown length
* 35790 36632: contig of 843 bp in length
* gap of unknown length
* 36633 37478: contig of 846 bp in length
* gap of unknown length
* 37479 38321: contig of 843 bp in length
* gap of unknown length
* 38322 39163: contig of 842 bp in length
* gap of unknown length
* 39164 40013: contig of 850 bp in length
* gap of unknown length
* 40014 40858: contig of 845 bp in length
* gap of unknown length
* 40859 41719: contig of 861 bp in length
* gap of unknown length
* 41720 42581: contig of 862 bp in length
* gap of unknown length
* 42582 43449: contig of 868 bp in length
* gap of unknown length
* 43450 44290: contig of 841 bp in length
* gap of unknown length
* 44291 45216: contig of 926 bp in length
* gap of unknown length
* 45217 46065: contig of 849 bp in length
* gap of unknown length
* 46066 46890: contig of 825 bp in length
* gap of unknown length
* 46891 47747: contig of 857 bp in length
* gap of unknown length
* 47748 48582: contig of 835 bp in length
* gap of unknown length
* 48583 49444: contig of 862 bp in length
* gap of unknown length
* 49445 50302: contig of 858 bp in length
* gap of unknown length
* 50303 51154: contig of 852 bp in length
* gap of unknown length
* 51155 52011: contig of 857 bp in length
* gap of unknown length
* 52012 52864: contig of 853 bp in length
* gap of unknown length
* 52865 53680: contig of 816 bp in length
* gap of unknown length

* 53681 54500: contig of 820 bp in length
 * gap of unknown length
 * 54501 55357: contig of 857 bp in length
 * gap of unknown length
 * 55358 56213: contig of 856 bp in length
 * gap of unknown length
 * 56214 57033: contig of 820 bp in length
 * gap of unknown length
 * 57034 58052: contig of 1019 bp in length
 * gap of unknown length
 * 58053 58884: contig of 832 bp in length
 * gap of unknown length
 * 58885 59738: contig of 854 bp in length
 * gap of unknown length
 * 59739 60599: contig of 861 bp in length
 * gap of unknown length
 * 60600 61442: contig of 843 bp in length
 * gap of unknown length
 * 61443 62284: contig of 842 bp in length
 * gap of unknown length
 * 62285 63113: contig of 829 bp in length
 * gap of unknown length
 * 63114 63949: contig of 836 bp in length
 * gap of unknown length
 * 63950 64788: contig of 839 bp in length
 * gap of unknown length
 * 64789 65709: contig of 921 bp in length
 * gap of unknown length
 * 65710 66552: contig of 843 bp in length
 * gap of unknown length

Query Match 72.4%; Score 36.2; DB 2; Length 73282;
 Best Local Similarity 82.0%; Pred. No. 0.0053;
 Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ACTTAAGCTTATAGGATGACCTGCGCGCTTCCAGTCGGGAAACCTGCG 50
 |||||
 Db 59421 ANTTAATTGCGTTCGCTCACTGCGCGCTTCCAGTCGGGAAACCTGCG 59470
 |||||

RESULT 7
 AC073998
 LOCUS 217141 bp DNA linear HTG 30-SEP-2000
 DEFINITION Homo sapiens clone RP11-337M23, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC073998
 VERSION AC073998.4 GI:10440718
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 217141)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-337M23
 Unpublished
 2 (bases 1 to 217141)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Melidrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (09-JUL-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Sep 30, 2000 this sequence version replaced gi:9838018.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7934
 Center clone name: 337_M_23

* NOTE: This record contains 270 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1 685: contig of 685 bp in length
 * 686 785: gap of 100 bp
 * 786 1466: contig of 681 bp in length
 * 1467 1566: gap of 100 bp
 * 1567 2244: contig of 678 bp in length
 * 2245 2345: gap of 100 bp
 * 2346 3018: contig of 674 bp in length
 * 3019 3118: gap of 100 bp
 * 3119 3779: contig of 661 bp in length
 * 3780 3879: gap of 100 bp
 * 3880 4565: contig of 686 bp in length
 * 4566 4665: gap of 100 bp
 * 4666 5345: contig of 680 bp in length
 * 5346 5445: gap of 100 bp
 * 5446 6122: contig of 677 bp in length
 * 6123 6222: gap of 100 bp
 * 6223 6891: contig of 669 bp in length
 * 6892 6991: gap of 100 bp
 * 6992 7674: contig of 683 bp in length
 * 7675 7774: gap of 100 bp
 * 7775 8458: contig of 684 bp in length
 * 8459 8558: gap of 100 bp
 * 8559 9249: contig of 691 bp in length
 * 9250 9349: gap of 100 bp
 * 9350 10031: contig of 682 bp in length
 * 10032 10131: gap of 100 bp
 * 10132 10831: contig of 700 bp in length
 * 10832 10932: gap of 100 bp
 * 10933 11617: contig of 686 bp in length
 * 11618 11717: gap of 100 bp
 * 11718 12390: contig of 673 bp in length
 * 12391 12490: gap of 100 bp
 * 12491 13169: contig of 679 bp in length
 * 13170 13269: gap of 100 bp
 * 13270 13947: contig of 678 bp in length
 * 13948 14047: gap of 100 bp
 * 14048 14714: contig of 667 bp in length
 * 14715 14814: gap of 100 bp
 * 14815 15505: contig of 691 bp in length
 * 15506 15605: gap of 100 bp
 * 15606 15284: contig of 679 bp in length
 * 16285 16384: gap of 100 bp

```
* 16385 17071: contig of 687 bp in length
* 17072 17171: gap of 100 bp
* 17172 17837: contig of 666 bp in length
* 17838 17937: gap of 100 bp
* 17938 18614: contig of 677 bp in length
* 18615 18714: gap of 100 bp
* 18715 19396: contig of 682 bp in length
* 19397 19496: gap of 100 bp
* 19497 20186: contig of 690 bp in length
* 20187 20286: gap of 100 bp
* 20287 20970: contig of 684 bp in length
* 20971 21070: gap of 100 bp
* 21071 21757: contig of 687 bp in length
* 21758 21857: gap of 100 bp
* 21858 22550: contig of 693 bp in length
* 22551 22650: gap of 100 bp
* 22651 23332: contig of 682 bp in length
* 23333 23432: gap of 100 bp
* 23433 24115: contig of 683 bp in length
* 24116 24215: gap of 100 bp
* 24216 24899: contig of 684 bp in length
* 24900 24999: gap of 100 bp
* 25000 25682: contig of 683 bp in length
* 25683 25782: gap of 100 bp
* 25783 26474: contig of 692 bp in length
* 26475 26574: gap of 100 bp
* 26575 27258: contig of 684 bp in length
* 27259 27358: gap of 100 bp
* 27359 28040: contig of 682 bp in length
* 28041 28140: gap of 100 bp
* 28141 28821: contig of 681 bp in length
* 28822 28921: gap of 100 bp
* 28922 29601: contig of 680 bp in length
* 29602 29701: gap of 100 bp
* 29702 30385: contig of 684 bp in length
* 30386 30485: gap of 100 bp
* 30486 31173: contig of 688 bp in length
* 31174 31273: gap of 100 bp
* 31274 31959: contig of 686 bp in length
* 31960 32059: gap of 100 bp
* 32060 32742: contig of 683 bp in length
* 32743 32842: gap of 100 bp
* 32843 33528: contig of 686 bp in length
* 33529 33628: gap of 100 bp
* 33629 34321: contig of 693 bp in length
* 34322 34421: gap of 100 bp
* 34422 35119: contig of 698 bp in length
* 35120 35219: gap of 100 bp
* 35220 35898: contig of 679 bp in length
* 35899 35998: gap of 100 bp
* 35999 36682: contig of 684 bp in length
* 36683 36782: gap of 100 bp
* 36783 37470: contig of 688 bp in length
* 37471 37570: gap of 100 bp
* 37571 38240: contig of 670 bp in length
* 38241 38340: gap of 100 bp
* 38341 39025: contig of 685 bp in length
* 39026 39125: gap of 100 bp
* 39126 39824: contig of 699 bp in length
* 39825 39924: gap of 100 bp
* 39925 40610: contig of 686 bp in length
* 40611 40710: gap of 100 bp
* 40711 41397: contig of 687 bp in length
* 41398 41497: gap of 100 bp
* 41498 42184: contig of 687 bp in length
* 42185 42284: gap of 100 bp
* 42285 42968: contig of 684 bp in length
* 42969 43068: gap of 100 bp
* 43069 43759: contig of 691 bp in length
* 43760 43859: gap of 100 bp
* 43860 44530: contig of 671 bp in length
* 44531 44630: gap of 100 bp
* 44631 45318: contig of 688 bp in length
```

```
* 45319 45418: gap of 100 bp
* 45419 46098: contig of 680 bp in length
* 46099 46198: gap of 100 bp
* 46199 46860: contig of 662 bp in length
* 46861 46960: gap of 100 bp
* 46961 47650: contig of 690 bp in length
* 47651 47750: gap of 100 bp
* 47751 48425: contig of 675 bp in length
* 48426 48525: gap of 100 bp
* 48526 49210: contig of 685 bp in length
* 49211 49310: gap of 100 bp
* 49311 49997: contig of 687 bp in length
* 49998 50097: gap of 100 bp
* 50098 50790: contig of 693 bp in length
* 50791 50890: gap of 100 bp
* 50891 51580: contig of 690 bp in length
* 51581 51680: gap of 100 bp
* 51681 52361: contig of 681 bp in length
* 52362 52461: gap of 100 bp
* 52462 53117: contig of 656 bp in length
* 53118 53217: gap of 100 bp
* 53218 53903: contig of 686 bp in length
* 53904 54003: gap of 100 bp

Query Match      72.4%  Score 36.2;  DB 2;  Length 217141;
Best Local Similarity 83.7%  Pred. No. 0.0054;
Matches 41;  Conservative 0;  Mismatches 8;  Indels 0;  Gaps 0;

Qy  2  CTTAAGCTTATAGCATGACTGCCGCTTTCCAGTCGGAAACCTGTGCG 50
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  162194  CTTAATTGCGTTCGCTCACTGCCGCTTTCCAGTCGGAAACCTGTGCG 162242

RESULT 8
G39061          G39061          604 bp  DNA  linear  STS 01-FEB-2001
LOCUS          Z11905 Zebrafish AB Danio rerio STS genomic, sequence tagged site.
ACCESSION      G39061
VERSION        G39061.1  GI:3358270
KEYWORDS       STS.
SOURCE         Danio rerio (zebrafish)
ORGANISM       Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE      1 (bases 1 to 604)
AUTHORS       Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S.,
               Jackson,D., de Sauvage,F., Jacob,H. and Fishman,M.C.
               Zebrafish genetic map with 2000 microsatellite markers
               Genomics 58 (3), 219-232 (1999)
TITLE          JOURNAL
MEDLINE        99303552
PUBMED         10373319

COMMENT
Contact: Mark C. Fishman
Cardiovascular Research Center
Massachusetts General Hospital
Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
Fax: 617/265806
Email: fishman@gh.cvr.c.harvard.edu
http://zebrafish.mgh.harvard.edu
Primer A: CACCGAGCTTCACTGACGTA
Primer B: ATACACACCAAGCGGACAT
STS size: 112
PCR Profile:
  Presoak:      94 degrees C for 5.0 minutes
  Denaturation: 94 degrees C for 1.0 minute
  Annealing:    58 degrees C for 1.0 minute
  Polymerization: 72 degrees C for 1.5 minute
  PCR Cycles:   27
  Thermal Cycler: MJ Research PTC-100
Protocol:
  Template:     10 ng
  Primer:       each 375 nM
```

dNTPs: each 200 uM
Taq Polymerase: 0.034 units/uI
Total Vol: 10 uI

Buffer: MgCl2: 1.5 mM
XCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

Primers are available from Research Genetics Inc.
(http://www.resgen.com phone: 800-533-4363).

FEATURES

source
1..604
/organism="Danio rerio"
/mol_type="genomic DNA"
/strain="AB"
/db_xref="taxon:7955"
/sex="P"
/clone_lib="zebrafish AB"
/dev_stage="Adult"
/lab_host="DH5alpha/IQ"
/note="Vector: m13mp19 with added BstXI site. V-type: Phage; Genomic DNA from a single adult Zebrafish of AB strain was digested with AluI, Cae8I, HaeIII, NlaVI, or RsaI. Fragments in the range of 250-500 bp were gel purified and a BstXI linker was added. The fragments were cloned into a modified M13mp19 vector and transformed into E. Coli DH5alpha. Microsatellite sequences were screened with labeled d(CA)15 and d(GT)15 oligonucleotide probes."
STS
188..299
primer_bind
188..207
complement(280..299)

ORIGIN

Query Match 71.6%; Score 35.8; DB 11; Length 604;
Best Local Similarity 83.3%; Pred. No. 0.0067;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 3 TTAACTTATACGATGACTCCCGCTTCCAGTCGGGAACCTCTCG 50
|||||
DB 515 TTAAATGTTGCGTCACTCCCGCTTCCAGTCGGGAACCTCTCG 562
|||||

RESULT 9

BD269301/c
LOCUS BD269301 745 bp DNA linear PAT 17-JUL-2003
DEFINITION 33 human secreted proteins.
ACCESSION BD269301
VERSION BD269301.1 GI:33079069
KEYWORDS JP 2002534972-A/36.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS Rosen,C.A., Ruben,S.M., Ebner,R., Young,P.E., Ni,J., Moore,P.A., Komatsoulis,G. and Birse,C.E.
TITLE Patent: JP 2002534972-A 36 22-OCT-2002;
JOURNAL HUMAN GENOME SCIENCES INC
COMMENT OS Homo sapiens (human)
PN JP 2002534972-A/36
PD 22-OCT-2002
PF 18-JAN-2000 JP 2000594904
PR 19-JAN-1999 US 60/116330
PI CRAIG A ROSEN, STEVEN M RUBEN, REINHARD EBNER, PAUL E YOUNG, JIAN NI,
PI PAUL A MOORE, GEORGE KOMATSIOULIS, CHARLES E BIRSE PC
C12N15/09,A61K31/7115,A61K35/76,A61K38/00,A61K45/00,A61K48/00, PC
A61P1/00, PC

A61P1/16,A61P3/10,A61P5/00,A61P7/02,A61P7/04,A61P7/06,A61P7/08, PC
A61P9/00,
PC A61P9/06,A61P9/08,A61P9/10,A61P9/12,A61P9/14,A61P11/00 PC
A61P11/06,A61P13/02,
PC A61P13/12,A61P15/00,A61P17/02,A61P17/06,A61P19/02,A61P21/00,
PC A61P21/04,
PC A61P25/00,A61P25/14,A61P25/16,A61P25/28,A61P25/30,A61P27/02,
PC A61P27/06,
PC A61P29/00,A61P29/00,A61P31/04,A61P31/10,A61P31/12,A61P31/18,
PC A61P33/00,
PC A61P35/00,A61P35/02,A61P35/04,A61P37/02,A61P37/08,A61P39/02,
PC A61P43/00,
PC C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
C12P21/02,C12Q1/02,
PC C12Q1/68,G01N33/15,G01N33/50,G01N33/53,G01N33/56,G01N33/68,
PC C12N15/00,
PC C12N5/00,A61K37/02
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
CC n equals a,t,g, or c
FH Key Location/Qualifiers
FT source 1..745 /organism='Homo sapiens (human)'.
FEATURES
source
1..745
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 71.6%; Score 35.8; DB 6; Length 745;
Best Local Similarity 83.3%; Pred. No. 0.0067;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 3 TTAACTTATACGATGACTCCCGCTTCCAGTCGGGAACCTCTCG 50
|||||
DB 168 TTAAATGTTGCGTCACTCCCGCTTCCAGTCGGGAACCTCTCG 121
|||||

RESULT 10

G40613
LOCUS G40613 826 bp DNA linear STS 01-FEB-2001
DEFINITION Z7956 Zebrafish AB Danio rerio STS genomic, sequence tagged site.
ACCESSION G40613
VERSION G40613.1 GI:3359822
KEYWORDS STS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 826)
AUTHORS Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S., Jackson,D., de Sauvage,F., Jacob,H. and Fishman,M.C.
TITLE Zebrafish genetic map with 2000 microsatellite markers
JOURNAL Genomics 58 (3), 219-232 (1999)
MEDLINE 99303552
PUBMED 10373319
COMMENT
Contact: Mark C. Fishman
Cardiovascular Research Center
Massachusetts General Hospital
Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
Fax: 6177265806
Email: fishman@gh.cvr.harvard.edu
http://zebrafish.mgh.harvard.edu
Primer A: CACATGTGACCGACTCTA

```

Primer B: TCTCTCCCTGGACATCATC
STS size: 142
PCR Profile:
  Presoak: 94 degrees C for 5.0 minutes
  Denaturation: 94 degrees C for 1.0 minute
  Annealing: 58 degrees C for 1.0 minute
  Polymerization: 72 degrees C for 1.5 minute
  PCR Cycles: 27
  Thermal Cycler: MJ Research PTC-100

Protocol:
  Template: 10 ng
  Primer: each 375 nM
  dNTPs: each 200 uM
  Tag Polymerase: 0.034 units/ul
  Total Vol: 10 ul

Buffer:
  MgCL2: 1.5 mM
  KCl: 50 mM
  Tris-HCl: 10 mM
  pH: 8.3

Primers are available from Research Genetics Inc.
(http://www.resgen.com phone: 800-533-4363).

FEATURES
  source
    1..826
    /organism="Danio rerio"
    /mol_type="genomic DNA"
    /strain="AB"
    /db_xref="taxon:7955"
    /map="LG 16"
    /sex="P"
    /clone_lib="Zebrafish AB"
    /dev_stage="Adult"
    /lab_host="DH5alphaF'IQ"
    /note="Vector: m13MP19 with added BstXI site; V-type:
    Phage; Genomic DNA from a single adult zebrafish of AB
    strain was digested with AluI, Cac8I, HaeIII, NlaVI, or
    RsaI. Fragments in the range of 250-500 bp were gel
    purified and a BstXI linker was added. The fragments were
    cloned into a modified M13mp19 vector and transformed
    into E. Coli DH5alpha. Microsatellite sequences were
    screened with labeled d(CA)15 and d(GT)15 oligonucleotide
    probes."
    STS
    primer_bind
    primer_bind
    complement(187..206)

ORIGIN
  Query Match 71.6%; Score 35.8; DB 11; Length 826;
  Best Local Similarity 83.3%; Pred. No. 0.0068;
  Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCGATGACTCCCGCTTCCAGTCGGGAACCTGTGC 50
    |||||
Db 516 TTAANTGCGTTCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGC 563

RESULT 11
LOCUS G40162 979 bp DNA linear STS 01-FEB-2001
DEFINITION Z20177 Zebrafish AB Danio rerio STS genomic, sequence tagged site.
ACCESSION G40162
VERSION G40162.1 GI:3359371
KEYWORDS STS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Danio.
REFERENCE
  1. (bases 1 to 979)
  Shimoda,N., Knapik,E.W., Ziniti,J., Sim,C., Yamada,E., Kaplan,S.,
  Jackson,D., de Sauvage,F., Jacob,H. and Fishman,M.C.

```

```

Zebrafish genetic map with 2000 microsatellite markers
Genomics 58 (3), 219-232 (1999)
9303552
10373119
PUBMED
COMMENT
  Contact: Mark C. Fishman
  Cardiovascular Research Center
  Massachusetts General Hospital
  Mail code 1494100A, 149 13th Street, Charlestown, MA 02129, USA
  Fax: 6177265806
  Email: fishman@mh.cvr.harvard.edu
  http://zebrafish.mgh.harvard.edu
  Primer A: TCAATCTGTCAAACTCCGCA
  Primer B: CGCTTATAGGCTGCAGAG
  STS size: 244
  PCR Profile:
    Presoak: 94 degrees C for 5.0 minutes
    Denaturation: 94 degrees C for 1.0 minute
    Annealing: 58 degrees C for 1.0 minute
    Polymerization: 72 degrees C for 1.5 minute
    PCR Cycles: 27
    Thermal Cycler: MJ Research PTC-100

  Protocol:
    Template: 10 ng
    Primer: each 375 nM
    dNTPs: each 200 uM
    Tag Polymerase: 0.034 units/ul
    Total Vol: 10 ul

  Buffer:
    MgCL2: 1.5 mM
    KCl: 50 mM
    Tris-HCl: 10 mM
    pH: 8.3

  Primers are available from Research Genetics Inc.
  (http://www.resgen.com phone: 800-533-4363).

FEATURES
  source
    1..979
    /organism="Danio rerio"
    /mol_type="genomic DNA"
    /strain="AB"
    /db_xref="taxon:7955"
    /map="LG 16"
    /sex="P"
    /clone_lib="Zebrafish AB"
    /dev_stage="Adult"
    /lab_host="DH5alphaF'IQ"
    /note="Vector: m13MP19 with added BstXI site; V-type:
    Phage; Genomic DNA from a single adult zebrafish of AB
    strain was digested with AluI, Cac8I, HaeIII, NlaVI, or
    RsaI. Fragments in the range of 250-500 bp were gel
    purified and a BstXI linker was added. The fragments were
    cloned into a modified M13mp19 vector and transformed
    into E. Coli DH5alpha. Microsatellite sequences were
    screened with labeled d(CA)15 and d(GT)15 oligonucleotide
    probes."
    STS
    primer_bind
    primer_bind
    complement(344..363)

ORIGIN
  Query Match 71.6%; Score 35.8; DB 11; Length 979;
  Best Local Similarity 83.3%; Pred. No. 0.0068;
  Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCGATGACTCCCGCTTCCAGTCGGGAACCTGTGC 50
    |||||
Db 726 TTAANTGCGTTCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGC 773

RESULT 12
AC023542

```

```

LOCUS       AC023542               76295 bp    DNA    linear    HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-165P21 map 15, LOW-PASS
SEQUENCE    AC023542
VERSION     AC023542.2   GI:9156023
KEYWORDS    HTG: HTGS PHASE0.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 76295)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 15, clone RP11-165P21
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 76295)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bada,F., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collinmore,A., Cooke,P.,
DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6978228.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6028
Center clone name: 165_P_21
-----
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1       788: contig of 788 bp in length
*      789      888: gap of 100 bp
*      889      1671: contig of 783 bp in length
*      1672      2563: contig of 792 bp in length
*      2564      2663: gap of 100 bp
*      2664      3454: contig of 791 bp in length
*      3455      3554: gap of 100 bp
*      3555      4331: contig of 777 bp in length
*      4332      4431: gap of 100 bp
*
* 4432      5258: contig of 827 bp in length
* 5259      5358: gap of 100 bp
* 5359      6154: contig of 796 bp in length
* 6155      6254: gap of 100 bp
* 6255      7042: contig of 788 bp in length
* 7043      7142: gap of 100 bp
* 7143      7919: contig of 777 bp in length
* 7920      8019: gap of 100 bp
* 8020      8788: contig of 769 bp in length
* 8789      9667: gap of 100 bp
* 9668      9667: contig of 779 bp in length
* 9668      9667: gap of 100 bp
* 9668      10566: contig of 799 bp in length
* 10567      10566: gap of 100 bp
* 10567      11444: contig of 778 bp in length
* 11445      11544: gap of 100 bp
* 11545      12321: contig of 777 bp in length
* 12322      12421: gap of 100 bp
* 12422      13200: contig of 779 bp in length
* 13201      13300: gap of 100 bp
* 13301      14086: contig of 786 bp in length
* 14087      14186: gap of 100 bp
* 14187      14984: contig of 798 bp in length
* 14985      15084: gap of 100 bp
* 15085      15846: contig of 762 bp in length
* 15847      15946: gap of 100 bp
* 15947      16726: contig of 780 bp in length
* 16727      16826: gap of 100 bp
* 16827      17586: contig of 760 bp in length
* 17587      17686: gap of 100 bp
* 17687      18461: contig of 775 bp in length
* 18462      18561: gap of 100 bp
* 18562      19340: contig of 779 bp in length
* 19341      19440: gap of 100 bp
* 19441      20236: contig of 796 bp in length
* 20237      20336: gap of 100 bp
* 20337      21119: contig of 783 bp in length
* 21120      21219: gap of 100 bp
* 21220      22015: contig of 796 bp in length
* 22016      22115: gap of 100 bp
* 22116      22319: contig of 804 bp in length
* 22920      23019: gap of 100 bp
* 23020      23816: contig of 797 bp in length
* 23817      23916: gap of 100 bp
* 23917      24688: contig of 772 bp in length
* 24689      24788: gap of 100 bp
* 24789      25596: contig of 808 bp in length
* 25597      25696: gap of 100 bp
* 25697      26481: contig of 785 bp in length
* 26482      26581: gap of 100 bp
* 26582      27368: contig of 787 bp in length
* 27369      27468: gap of 100 bp
* 27469      28245: contig of 777 bp in length
* 28246      28345: gap of 100 bp
* 28346      29118: contig of 773 bp in length
* 29119      29218: gap of 100 bp
* 29219      30006: contig of 788 bp in length
* 30007      30106: gap of 100 bp
* 30107      30897: contig of 791 bp in length
* 30898      30997: gap of 100 bp
* 30998      31778: contig of 781 bp in length
* 31779      31878: gap of 100 bp
* 31879      32652: contig of 774 bp in length
* 32653      32752: gap of 100 bp
* 32753      33532: contig of 780 bp in length
* 33533      33632: gap of 100 bp
* 33633      34408: contig of 776 bp in length
* 34409      34508: gap of 100 bp
* 34509      35276: contig of 768 bp in length
* 35277      35376: gap of 100 bp
* 35377      36155: contig of 779 bp in length
* 36156      36255: gap of 100 bp
* 36256      37048: contig of 793 bp in length

```

```
* 37049 37148: gap of 100 bp
* 37149 37949: contig of 801 bp in length
* 37950 38049: gap of 100 bp
* 38050 38830: contig of 781 bp in length
* 38831 38930: gap of 100 bp
* 38931 39716: contig of 786 bp in length
* 39717 39816: gap of 100 bp
* 39817 40607: contig of 791 bp in length
* 40608 40707: gap of 100 bp
* 40708 41499: contig of 792 bp in length
* 41500 41599: gap of 100 bp
* 41600 42358: contig of 759 bp in length
* 42359 42458: gap of 100 bp
* 42459 43225: contig of 767 bp in length
* 43226 43325: gap of 100 bp
* 43326 44104: contig of 779 bp in length
* 44105 44204: gap of 100 bp
* 44205 44988: contig of 784 bp in length
* 44989 45088: gap of 100 bp
* 45089 45887: contig of 799 bp in length
* 45888 45987: gap of 100 bp
* 45988 46764: contig of 777 bp in length
* 46765 46864: gap of 100 bp
* 46865 47646: contig of 782 bp in length
* 47647 47746: gap of 100 bp
* 47747 48523: contig of 777 bp in length
* 48524 48623: gap of 100 bp
* 48624 49404: contig of 781 bp in length
* 49405 49504: gap of 100 bp
* 49505 50294: contig of 790 bp in length
* 50295 50394: gap of 100 bp
* 50395 51164: contig of 770 bp in length
* 51165 51264: gap of 100 bp
* 51265 52052: contig of 788 bp in length
* 52053 52152: gap of 100 bp
* 52153 52941: contig of 789 bp in length
* 52942 53041: gap of 100 bp
* 53042 53819: contig of 778 bp in length
* 53820 53919: gap of 100 bp
* 53920 54694: contig of 775 bp in length
* 54695 54794: gap of 100 bp
* 54795 55560: contig of 766 bp in length
* 55561 55660: gap of 100 bp
* 55661 56450: contig of 790 bp in length
* 56451 56550: gap of 100 bp
* 56551 57325: contig of 775 bp in length
* 57326 57425: gap of 100 bp
* 57426 58212: contig of 787 bp in length
* 58213 58312: gap of 100 bp
* 58313 59110: contig of 798 bp in length
* 59111 59210: gap of 100 bp
* 59211 59992: contig of 782 bp in length
* 59993 60092: gap of 100 bp
* 60093 60899: contig of 807 bp in length
```

```
Query Match 71.6%; Score 35.8; DB 2; Length 76295;
Best Local Similarity 83.3%; Pred. No. 0.0076;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

```
Qy 3 TTAAGCTTATAGCGATGACTGCCCGCTTCCAGTCGGGAAACCTGTCG 50
|||||
Db 25290 TTAATGCTTGGCGTCACTGCCCGCTTCCAGTCGGGAAACCTGTCG 25337
```

```
RESULT 13
CQ406071 LOCUS CQ406071 586 bp DNA linear PAT 23-JAN-2004
DEFINITION Sequence 13142 from Patent WO0170979.
ACCESSION CQ406071
VERSION CQ406071.1 GI:41313852
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS Lee, J. and Lillie, J.
TITLE Genes, compositions, kits, and method for identification,
assessment, prevention, and therapy of ovarian cancer
JOURNAL Patent: WO 0170979-A 13142 27-SEP-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source Location/Qualifiers
1 .586
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
```

ORIGIN

```
Query Match 71.2%; Score 35.6; DB 6; Length 586;
Best Local Similarity 82.0%; Pred. No. 0.008;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy 1 ACITTAAGCTTATAGCGATGACTGCCCGCTTCCAGTCGGGAAACCTGTCG 50
|||||
Db 516 ATTTAATGCTTGGCGTCACTGCCCGCTTCCAGTCGGGAAACCTGTCG 565
```

RESULT 14

```
AC023384 75002 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 11 clone RP11-589112 map 11, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC023384
VERSION AC023384.2 GI:9144035
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-589112
JOURNAL Unpublished
```

```
REFERENCE
AUTHORS 2 (bases 1 to 75002)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepei, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
DeArellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
Fenster, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Landers, T., LARGOCQUE, K., Lenoczky, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Miranda, C., Mlenda, V., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T. M.,
Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
Riley, R. K., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.
Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6970532.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A. F. A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
```

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5114
Center clone name: 589_I_12

* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 724: contig of 724 bp in length
* 725 824: gap of 100 bp
* 825 1573: contig of 749 bp in length
* 1574 1673: gap of 100 bp
* 1674 2455: contig of 782 bp in length
* 2456 2555: gap of 100 bp
* 2556 3328: contig of 773 bp in length
* 3329 3428: gap of 100 bp
* 3429 4179: contig of 751 bp in length
* 4180 4279: gap of 100 bp
* 4280 5022: contig of 743 bp in length
* 5023 5122: gap of 100 bp
* 5123 5865: contig of 743 bp in length
* 5866 6703: contig of 738 bp in length
* 6704 6803: gap of 100 bp
* 6804 7554: contig of 751 bp in length
* 7555 8408: contig of 754 bp in length
* 8409 8508: gap of 100 bp
* 8509 9254: contig of 746 bp in length
* 9255 9354: gap of 100 bp
* 9355 10118: contig of 764 bp in length
* 10119 10218: gap of 100 bp
* 10219 10978: contig of 760 bp in length
* 10979 11078: gap of 100 bp
* 11079 11850: contig of 772 bp in length
* 11851 11950: gap of 100 bp
* 11951 12712: contig of 762 bp in length
* 12713 12812: gap of 100 bp
* 12813 13553: contig of 741 bp in length
* 13554 13553: gap of 100 bp
* 13554 14392: contig of 739 bp in length
* 14393 14492: gap of 100 bp
* 14493 15244: contig of 752 bp in length
* 15245 15344: gap of 100 bp
* 15345 16095: contig of 751 bp in length
* 16096 16195: gap of 100 bp
* 16196 16927: contig of 732 bp in length
* 16928 17027: gap of 100 bp
* 17028 17766: contig of 739 bp in length
* 17767 17866: gap of 100 bp
* 17867 18115: contig of 749 bp in length
* 18116 18715: gap of 100 bp
* 18716 19489: contig of 774 bp in length
* 19490 19589: gap of 100 bp
* 19590 20344: contig of 755 bp in length
* 20345 20444: gap of 100 bp
* 20445 21205: contig of 761 bp in length
* 21206 21305: gap of 100 bp
* 21306 22089: contig of 784 bp in length
* 22090 22189: gap of 100 bp
* 22190 22972: contig of 783 bp in length
* 22973 23072: gap of 100 bp
* 23073 23826: contig of 754 bp in length
* 23827 23526: gap of 100 bp
* 23927 24674: contig of 748 bp in length

* 24675 24774: gap of 100 bp
* 24775 25507: contig of 733 bp in length
* 25508 25607: gap of 100 bp
* 25608 26347: contig of 740 bp in length
* 26348 26447: gap of 100 bp
* 26448 27202: contig of 755 bp in length
* 27203 27302: gap of 100 bp
* 27303 28052: contig of 750 bp in length
* 28053 28152: gap of 100 bp
* 28153 28909: contig of 757 bp in length
* 28910 29009: gap of 100 bp
* 29010 29741: contig of 732 bp in length
* 29742 29841: gap of 100 bp
* 29842 30624: contig of 783 bp in length
* 30625 30724: gap of 100 bp
* 30725 31497: contig of 773 bp in length
* 31498 31597: gap of 100 bp
* 31598 32373: contig of 776 bp in length
* 32374 32473: gap of 100 bp
* 32474 33222: contig of 749 bp in length
* 33223 33322: gap of 100 bp
* 33323 34072: contig of 750 bp in length
* 34073 34172: gap of 100 bp
* 34173 34932: contig of 760 bp in length
* 34933 35032: gap of 100 bp
* 35033 35791: contig of 759 bp in length
* 35792 35891: gap of 100 bp
* 35892 36636: contig of 745 bp in length
* 36637 36736: gap of 100 bp
* 36737 37468: contig of 732 bp in length
* 37469 37568: gap of 100 bp
* 37569 38328: contig of 760 bp in length
* 38329 38428: gap of 100 bp
* 38429 39191: contig of 763 bp in length
* 39192 39291: gap of 100 bp
* 39292 40036: contig of 745 bp in length
* 40037 40136: gap of 100 bp
* 40137 40895: contig of 759 bp in length
* 40896 40995: gap of 100 bp
* 40996 41729: contig of 734 bp in length
* 41730 41829: gap of 100 bp
* 41830 42577: contig of 748 bp in length
* 42578 42677: gap of 100 bp
* 42678 43414: contig of 737 bp in length
* 43415 43514: gap of 100 bp
* 43515 44267: contig of 753 bp in length
* 44268 44367: gap of 100 bp
* 44368 45123: contig of 756 bp in length
* 45124 45223: gap of 100 bp
* 45224 45973: contig of 750 bp in length
* 45974 46073: gap of 100 bp
* 46074 46816: contig of 743 bp in length
* 46817 46916: gap of 100 bp
* 46917 47681: contig of 765 bp in length
* 47682 47781: gap of 100 bp
* 47782 48544: contig of 763 bp in length
* 48545 48644: gap of 100 bp
* 48645 49410: contig of 766 bp in length
* 49411 49510: gap of 100 bp
* 49511 50247: contig of 737 bp in length
* 50248 50347: gap of 100 bp
* 50348 51067: contig of 720 bp in length
* 51068 51167: gap of 100 bp
* 51168 51915: contig of 748 bp in length
* 51916 52015: gap of 100 bp
* 52016 52745: contig of 730 bp in length
* 52746 52845: gap of 100 bp
* 52846 53584: contig of 739 bp in length
* 53585 53684: gap of 100 bp
* 53685 54427: contig of 743 bp in length
* 54428 54527: gap of 100 bp
* 54528 55292: contig of 765 bp in length
* 55293 55392: gap of 100 bp

```

* 55393 56164: contig of 772 bp in length
* 56165 56264: gap of 100 bp
* 56265 56985: contig of 721 bp in length
* 56986 57085: gap of 100 bp
* 57086 57829: contig of 744 bp in length
* 57830 57929: gap of 100 bp
* 57930 58677: contig of 748 bp in length

Query Match 71.2%; Score 35.6; DB 2; Length 75002;
Best Local Similarity 82.0%; Pred. No. 0.009; Indels 0; Gaps 0;
Matches 41; Conservative 0; Mismatches 9;

Qy 1 ACTTAAGCTTATAGCATGACTGCCGCTTTCCAGTCGGGAACCTGTGC 50
    |||||
Db 25456 AATTAAATGCTTGGCTCATGCGCGCTTTCCAGTCGGGAACCTGTGC 25505
    |||||

RESULT 15
AC016798 95127 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-1J15, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC016798
ACCESSION AC016798
VERSION AC016798.2 GI:9119829
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 95127)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-1J15
Unpublished
2 (bases 1 to 95127)
Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McSwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6532117.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2499
Center clone name: 1_J_15
-----
* NOTE: This record contains 97 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

```

```

* be preserved.
* 1
* 904 903: contig of 903 bp in length
* 1004 1003: gap of 100 bp
* 1852 1852: contig of 849 bp in length
* 1853 1852: gap of 100 bp
* 1953 2797: contig of 845 bp in length
* 2798 2897: gap of 100 bp
* 2898 3777: contig of 880 bp in length
* 3778 3877: gap of 100 bp
* 3878 4737: contig of 860 bp in length
* 4738 4837: gap of 100 bp
* 4838 5670: contig of 833 bp in length
* 5671 5770: gap of 100 bp
* 5771 6639: contig of 869 bp in length
* 6640 6739: gap of 100 bp
* 6740 7619: contig of 880 bp in length
* 7620 7719: gap of 100 bp
* 7720 8600: contig of 881 bp in length
* 8601 8700: gap of 100 bp
* 8701 9570: contig of 870 bp in length
* 9571 9670: gap of 100 bp
* 9671 10521: contig of 851 bp in length
* 10522 10621: gap of 100 bp
* 10622 11452: contig of 831 bp in length
* 11453 11552: gap of 100 bp
* 11553 12420: contig of 868 bp in length
* 12421 12520: gap of 100 bp
* 12521 13390: contig of 870 bp in length
* 13391 13490: gap of 100 bp
* 13491 14361: contig of 871 bp in length
* 14362 14461: gap of 100 bp
* 14462 15330: contig of 869 bp in length
* 15331 15430: gap of 100 bp
* 15431 16330: contig of 900 bp in length
* 16331 16430: gap of 100 bp
* 16431 17312: contig of 882 bp in length
* 17313 17412: gap of 100 bp
* 17413 18298: contig of 886 bp in length
* 18299 19243: contig of 845 bp in length
* 19244 19343: gap of 100 bp
* 19344 20190: contig of 847 bp in length
* 20191 20290: gap of 100 bp
* 20291 21162: contig of 872 bp in length
* 21163 21262: gap of 100 bp
* 21263 22130: contig of 868 bp in length
* 22131 22230: gap of 100 bp
* 22231 23094: contig of 864 bp in length
* 23095 23194: gap of 100 bp
* 23195 24072: contig of 878 bp in length
* 24073 24172: gap of 100 bp
* 24173 25016: contig of 844 bp in length
* 25017 25116: gap of 100 bp
* 25117 25993: contig of 877 bp in length
* 25994 26093: gap of 100 bp
* 26094 26980: contig of 887 bp in length
* 26981 27080: gap of 100 bp
* 27081 27909: contig of 829 bp in length
* 27910 28009: gap of 100 bp
* 28010 28886: contig of 877 bp in length
* 28887 29834: gap of 100 bp
* 29835 29934: contig of 848 bp in length
* 29936 30798: gap of 100 bp
* 30799 30898: contig of 864 bp in length
* 30899 31768: gap of 100 bp
* 31769 31868: contig of 897 bp in length
* 31869 32765: contig of 897 bp in length
* 32766 32865: gap of 100 bp
* 32866 33665: contig of 801 bp in length
* 33667 33766: gap of 100 bp
* 33767 34665: contig of 899 bp in length
* 34666 34765: gap of 100 bp

```


* 34766 35590: contig of 825 bp in length
* 35591 gap of 100 bp
* 35628: contig of 838 bp in length
* 36289: gap of 100 bp
* 36299: contig of 861 bp in length
* 37490 37589: gap of 100 bp
* 37590 38468: contig of 879 bp in length
* 38469 38568: gap of 100 bp
* 38569 39446: contig of 878 bp in length
* 39447 39546: gap of 100 bp
* 39547 40372: contig of 826 bp in length
* 40373 40472: gap of 100 bp
* 40473 41361: contig of 889 bp in length
* 41362 41461: gap of 100 bp
* 41462 42333: contig of 872 bp in length
* 42334 42433: gap of 100 bp
* 42434 43113: contig of 880 bp in length
* 43114 43413: gap of 100 bp
* 43414 44242: contig of 829 bp in length
* 44243 44342: gap of 100 bp
* 44343 45211: contig of 869 bp in length
* 45212 45311: gap of 100 bp
* 45312 46183: contig of 872 bp in length
* 46184 46283: gap of 100 bp
* 46284 47132: contig of 849 bp in length
* 47133 47232: gap of 100 bp
* 47233 48077: contig of 845 bp in length
* 48078 48177: gap of 100 bp
* 48178 49045: contig of 868 bp in length
* 49046 49145: gap of 100 bp
* 49146 50014: contig of 869 bp in length
* 50015 50114: gap of 100 bp
* 50115 50999: contig of 885 bp in length
* 51000 51099: gap of 100 bp
* 51099 51988: contig of 889 bp in length
* 51989 52088: gap of 100 bp
* 52089 52974: contig of 886 bp in length
* 52975 53074: gap of 100 bp
* 53075 53918: contig of 844 bp in length
* 53919 54018: gap of 100 bp
* 54019 54871: contig of 853 bp in length
* 54872 54971: gap of 100 bp
* 54972 55842: contig of 871 bp in length
* 55843 55942: gap of 100 bp
* 55943 56817: contig of 875 bp in length
* 56818 56917: gap of 100 bp
* 56918 57793: contig of 876 bp in length
* 57794 57893: gap of 100 bp
* 57894 58764: contig of 871 bp in length
* 58765 58864: gap of 100 bp
* 58865 59725: contig of 861 bp in length
* 59726 59825: gap of 100 bp
* 59826 60677: contig of 852 bp in length
* 60678 60777: gap of 100 bp
* 60778 61648: contig of 871 bp in length
* 61649 61748: gap of 100 bp
* 61749 62632: contig of 884 bp in length
* 62633 62732: gap of 100 bp
* 62733 63616: contig of 884 bp in length
* 63617 63716: gap of 100 bp
* 63717 64567: contig of 851 bp in length
* 64568 64667: gap of 100 bp
* 64668 65533: contig of 866 bp in length
* 65534 65633: gap of 100 bp
* 65634 66505: contig of 872 bp in length
* 66506 66605: gap of 100 bp
* 66606 67460: contig of 855 bp in length
* 67461 67560: gap of 100 bp
* 67561 68417: contig of 857 bp in length
* 68418 68517: gap of 100 bp
* 68518 69384: contig of 877 bp in length
* 69395 69494: gap of 100 bp
* 69495 70375: contig of 881 bp in length

Query Match 71.2%; Score 35.6; DB 2; Length 95127;
Best Local Similarity 82.0%; Pred. No. 0.0091;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 20592 ATTTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 20641

Search completed: October 24, 2005, 20:18:17
Job time : 311.908 secs

This Page Blank (uspto)

GenCore version 5.1.6

Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:41:23 ; Search time 360.936 Seconds
(without alignments)
5273.001 Million cell updates/sec

Title: US-09-896-888A-10

Perfect score: 50

Sequence: 1 acttaagtttagcgatga.....tcacgtcggaacctgtcg 50

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hc:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.4	76.8	412	9	CNS07H0H
2	36.8	73.6	125	8	AQ080570
3	36.8	73.6	174	8	AQ076128
4	36.8	73.6	177	8	AQ076155
5	36.8	73.6	311	8	AQ076114
6	36.8	73.6	372	7	CO990192
7	36.8	73.6	373	7	CO947892
8	36.8	73.6	425	2	BE403103
9	36.8	73.6	451	2	BE403103
10	36.8	73.6	455	9	CNS07H85
11	36.8	73.6	591	7	CO048749
12	36.8	73.6	641	9	AG068743
13	36.8	73.6	655	9	AG055437
14	36.8	73.6	681	9	AG125607
15	36.8	73.6	682	9	AG118925
16	36.8	73.6	723	9	AG068231
17	36.8	73.6	745	9	AG030401
18	36.8	73.6	755	9	AG102051
19	36.8	73.6	784	9	AG383436
20	36.8	73.6	885	2	BE229292
21	36.4	72.8	350	1	AU233787
22	36.4	72.8	357	1	AU278188
23	36.2	72.4	674	8	BH243517
24	35.8	71.6	232	1	AU069015

25	35.8	71.6	331	4	BJ691809
26	35.8	71.6	423	4	BJ668962
27	35.8	71.6	459	1	AU069599
28	35.8	71.6	477	5	BP874955
29	35.8	71.6	508	4	BJ671717
30	35.8	71.6	509	5	BQ907999
31	35.8	71.6	571	4	BM419897
32	35.8	71.6	586	6	CB865576
33	35.8	71.6	675	9	AG060720
34	35.8	71.6	699	9	AG043113
35	35.8	71.6	746	9	AG134861
36	35.8	71.6	789	9	ATH517143
37	35.8	71.6	817	7	CF569104
38	35.6	71.2	368	4	BI937595
39	35.6	71.2	410	8	B30742
40	35.6	71.2	496	2	BF703023
41	35.6	71.2	692	9	AG109195
42	35.6	71.2	719	9	AG045821
43	35.4	70.8	939	9	CU096240
44	35.2	70.4	117	8	AQ041632
45	35.2	70.4	120	9	CC961712

ALIGNMENTS

RESULT 1
CNS07H0H
LOCUS
DEFINITION
Anopheles gambiae CSS T7 end of clone 23p13 of library Notredame1
from strain FST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION
AL610451
VERSION
AL610451.1 GI:15916636
KEYWORDS
GSS.
SOURCE
Anopheles gambiae (African malaria mosquito)
ORGANISM
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE
1 (bases 1 to 412)
Genoscope.
Direct Submission
Submitted (01-OCT-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 412)
Roth,C.W., Brey,P.T., Ke,Z. and Collins,F.H.
Direct Submission
Submitted (01-OCT-2001) BMMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
Location/Qualifiers
1..412
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PES1"
/db_xref="taxon:7165"
/clone_lib="23p13"
/clone_lib="Notredame1"
/note="end : T7"

ORIGIN
Query Match 76.8%; Score 38.4; DB 9; Length 412;
Best Local Similarity 87.5%; Pred. No. 0.00023;
Matches 42; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 3 TTAGCTTATAGCGATGACTCCCGCTTTCAGTCGGGAACCTCTCG 50
DB 152 TTAATTTTTCGCTCACTCCCGCTTTCAGTCGGGAACCTCTCG 199

RESULT 2	125 bp	DNA	linear	GSS 20-AUG-1998		
AQ080570	CIT-HSP-2358M2.TF	CIT-HSP Homo sapiens	genomic clone	2358M2,		
LOCUS	genomic survey sequence.					
DEFINITION						
ACCESSION	AQ080570					
VERSION	AQ080570.1	GI:3441754				
KEYWORDS	GSS.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 125)					
TITLE	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.					
JOURNAL	Use of a random human BAC End Sequence Database for Sequence-Ready Map Building					
COMMENT	Unpublished (1998) Other_GSSs: CIT-HSP-2358M2.TR Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html. Seq primer: M13-21 Class: BAC ends.					
FEATURES	Location/Qualifiers					
source	1..125					
	/organism="Homo sapiens"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:9606"					
	/clone="2358M2"					
	/sex="Male"					
	/cell_type="Sperm"					
	/clone_lib="CIT-HSP"					
	/notes="vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"					
ORIGIN						
Query Match	73.6%	Score 36.8;	DB 8;	Length 125;		
Best Local Similarity	85.4%	Pred.No. 0.00082;				
Matches	41;	Conservative	0;	Mismatches 7; Indels 0; Gaps 0;		
Qy	3	TTAAGCTTATAGCGATCACTGCCGCTTTCAGTCGGAAACCTGTGCG	50			
Db	25	TTAATTCGTAGCGCTCACTGCCGCTTTCAGTCGGAAACCTGTGCG	72			
RESULT 3	AQ076128	174 bp	DNA	linear		
LOCUS	CIT-HSP-2364K4.TF	CIT-HSP Homo sapiens	genomic clone	2364K4,		
DEFINITION	genomic survey sequence.					
ACCESSION	AQ076128					
VERSION	AQ076128.1	GI:3436262				
KEYWORDS	GSS.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 174)					
TITLE	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.					
	Use of a random human BAC End Sequence Database for Sequence-Ready					

```

/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelOAC11; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      73.6%; Score 36.8; DB 8; Length 177;
Best Local Similarity 85.4%; Pred. No. 0.00086;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      3  TTAAGCTTATACGATGACTGCCCGCTTCCAGTCGGGAACCTGTCG 50
Db      83  TTAATAGCGTACGCTCACTGCCCGCTTCCAGTCGGGAACCTGTCG 130

RESULT 5
LOCUS      AQ076114
DEFINITION CIT-HSP-2368K7.TF CIT-HSP Homo sapiens genomic clone 2368K7,
genomic survey sequence.
ACCESSION  AQ076114
VERSION     AQ076114
KEYWORDS    GSS.
SOURCE      AQ076114.1 GI:3436248
            Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 311)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
TITLE       Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
JOURNAL     Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tcdb/hungen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.

FEATURES             Location/Qualifiers
     1..311
         /organism="Homo sapiens"
         /mol_type="genomic DNA"
         /db_xref="taxon:9606"
         /clone="2368K7"
         /sex="Male"
         /cell_type="Sperm"
         /clone_lib="CIT-HSP"
         /note="Vector: pBelOAC11; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      73.6%; Score 36.8; DB 8; Length 311;
Best Local Similarity 85.4%; Pred. No. 0.00095;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      3  TTAAGCTTATACGATGACTGCCCGCTTCCAGTCGGGAACCTGTCG 50
Db      232  TTAATAGCGTACGCTCACTGCCCGCTTCCAGTCGGGAACCTGTCG 279

RESULT 6
LOCUS      CO990192
DEFINITION UMC-pd3ov2-002-g02 Oviduct gilt D3 of estrous cycle pd3ov Sus

scrofa cDNA 3', mRNA sequence.
CO990192
CO990192.1 GI:51349466
EST.
KEYWORDS
SOURCE      Sus scrofa (pig)
ORGANISM    Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
            1  (bases 1 to 372)
            Kim,J., Whitworth,K.M., Springer,G.K., Bivens,N.J., Ries,J.E.,
            Woods,R.J., Spollen,W.G., Forrester,L.J., Mathialagan,N.,
            Prather,R.S. and Green,J.A.
            Large-scale Generation and Analysis of Expressed Sequence Tags from
            Porcine endometrium and oviduct
            Unpublished (2004)
            Contact: DNA Core Facility (Swine Project)
            Animal Science - RS Prather
            University of Missouri-Columbia
            M616 Medical Sciences Bldg., Columbia, MO 65212, USA
            Tel: (573)882-0428
            Fax: (573)884-5552
            Email: porcine@net.missouri.edu
            POLYA=No.

FEATURES             Location/Qualifiers
     1..372
         /organism="Sus scrofa"
         /mol_type="mRNA"
         /db_xref="taxon:9823"
         /dev_stage="Oviduct from a gilt on day 3 of the estrous
         cycle"
         /clone_lib="pd3ov"
         /note="Vector: pSP0rt1; Funding: A grant from the Monsanto
         Company to the University of Missouri. Genetic Source:
         Endometrium and oviduct tissues from various stages of the
         estrous cycle were collected from crossbred pigs (Sus
         scrofa domestica), frozen in liquid nitrogen immediately
         after collection, and stored at -80 degrees Celsius until
         RNA extraction. The specific tissues collected were Day 0
         and Day 3 whole oviducts and Days 3, 6, 10 and 12-14
         endometrium. More information regarding the methods can be
         found at:
         http://genome.rnet.missouri.edu/Swine/Methods.html.
         Library Construction (Standard Protocol): All procedures
         discussed in this section have been described in detail
         elsewhere (Soares et al., 1994; Bonaldo et al., 1996;
         Jiang et al., 2001). Total cellular RNA from each sample
         was isolated by using STAT-60 reagent (Tel-test,
         Friendswood, TX) and poly(A)+ RNA was obtained by two
         rounds of purification with the Oligotex mRNA isolation
         kit (Qiagen) according to the manufacturer's instructions.
         The oviduct libraries and the Day 3, 6 and 10 endometrium
         libraries were constructed essentially as described by the
         manufacturer's instructions provided with the SuperScript
         Plasmid System (Invitrogen, cat. no. 18248-013). Briefly,
         1mg of poly(A)+ RNA will be annealed at 37 degrees Celsius
         with 10mg of NotI-tag-dT18 oligonucleotide
         (GCTGCTCGCGCGC-tag-T18) and reverse transcribed at 37
         degrees Celsius with SuperScript II (Invitrogen) reverse
         transcriptase (Jiang et al., 2001). The 'tag' represents a
         tissue/stage-specific ten-base sequence identifier
         (http://genome.uiowa.edu/pubsoft/software.html) present in
         the oligonucleotide used to prime first-strand synthesis.
         Second strand synthesis was performed with T4 DNA
         polymerase in the presence of DNA ligase and RNase H.
         After second strand synthesis, the double-stranded cDNAs
         was ligated to SalI adapters (Invitrogen-Life
         Technologies) and digested with NotI. The cDNAs will be
         size selected by passage through cDNA size fractionation
         columns (Invitrogen-Life technologies). The cDNAs derived
         from each developmental stage of a particular tissue were
         mixed on an equimolar basis and ligated directionally into
         the NotI and SalI sites of the pSP0rt1 vector
         (Invitrogen). After ligation of the inserts, the plasmids

```


2. Jiang H, Bivens NJ, Ries JE, Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion BA, Mathialagan N, Prather RS, Lucy MC (2001) Constructing cDNA libraries with fewer clones that contain long poly(dA) tails. Biotechniques 31:38-42. 3. Soares MB, MF Banaldo, P Jelene, L Su, L Lawton, A Estrandiadis. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci, 91:9228-9232. TAG_TISSUE=Imm ovarian follicle TAG_SEQ=Not found"

ORIGIN

Query Match 73.6%; Score 36.8; DB 7; Length 373;
Best Local Similarity 85.4%; Pred. No. 0.00098;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCGTACTGCCGCTTTCCAGTCGGGAACCTGTGCG 50
|||||
Db 219 TTAATTGCTAGCGCTACTGCCGCTTTCCAGTCGGGAACCTGTGCG 266
|||||

RESULT 8
LOCUS C0954011
DEFINITION UMC-pnata14-004-e04 Perinatal ovary pnatal Sus scrofa cDNA 3', mRNA
sequence.
ACCESSION C0954011
VERSION C0954011.1 GI:51331074
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 425)
AUTHORS Jiang H., Whitworth K.M., Bivens N.J., Ries J.E., Woods R.J., Forrester L.J., Springer G.K., Mathialagan N., Agca C., Prather R.S. and Lucy M.C.
TITLE Large-scale Generation and Analysis of Expressed Sequence Tags from Porcine Ovary
JOURNAL Unpublished (2004)
COMMENT Contact: DNA Core Facility (Swine Project)
Animal Science - RS Prather
University of Missouri-Columbia
M616 Medical Sciences Bldg., Columbia, MO 65212, USA
Tel: (573)882-0428
Fax: (573)884-5552
Email: porcine@net.missouri.edu
POLYA=No.

FEATURES
source
1..425
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/dev_stage="perinatal ovary"
/clone_lib="pnatal"
/note="Vector: pSPORT1; Funding: A grant from the Monsanto Company to the University of Missouri. Genetic Source: Ovarian tissue (whole ovary, dissected follicles, or corpora lutea) was collected from crossbred pigs (Sus scrofa domestica), frozen in liquid nitrogen shortly after collection, and stored at -80 degrees Celsius until RNA extraction. The tissue from several individual pigs was pooled for the purpose of RNA extraction. The specific tissues collected were fetal whole ovary; neonatal whole ovary; prepubertal whole ovary; 2, 4, 6 and 8 mm growing follicles; Day 0 follicles; Day 5 small antral follicles and corpora lutea; Day 12 corpora lutea and Day 12 follicles. More information regarding the methods can be found at:
http://genome.rnet.missouri.edu/Swine/Methods.html.
Library Construction (Standard Protocol): All procedures discussed in this section have been described in detail elsewhere (Soares et al., 1994; Banaldo et al., 1996;

Jiang et al., 2001). Total cellular RNA from each sample was isolated by using STAT-60 reagent (Tel-Test, Friendswood, TX) and the poly(A)+ RNA was obtained by two rounds of purification with the Oligotex mRNA isolation kit (Qiagen) according to the manufacturer's instructions. The libraries were constructed essentially as described by the manufacturer's instructions provided with the SuperScript Plasmid System (Invitrogen, cat. no. 18248-013). Briefly, 1mg of poly(A)+ RNA will be annealed at 37 degrees Celsius with 10mg of NotI-tag-dT18 oligonucleotide (GCTCTCGCGCGC-tag-T18) and reverse transcribed at 37 degrees Celsius with SuperScript II (Invitrogen) reverse transcriptase (Jiang et al., 2001). The 'tag' represents a tissue/stage-specific ten-base sequence identifier (<http://genome.uiowa.edu/pubsoft/software.html>) present in the oligonucleotide used to prime first-strand synthesis. Second strand synthesis was performed with T4 DNA polymerase in the presence of DNA ligase and RNase H. After second strand synthesis, the double-stranded cDNAs was ligated to SalI adapters (Invitrogen-Life Technologies) and digested with NotI. The cDNAs will be size selected by passage through cDNA size fractionation columns (Invitrogen-Life technologies). The cDNAs derived from each developmental stage of a particular tissue were mixed on an equimolar basis and ligated directionally into the NotI and SalI sites of the pSPORT1 vector (Invitrogen). After ligation of the inserts, the plasmids will be electroporated into DH10B bacteria. Preliminary library characterization: Randomly chosen clones from each determine average insert size (96 clones) and by sequencing (~4 96-well plates) to confirm library quality (e.g. the presence of short polyA+ tails, genomic DNA contamination (must be <1%), ribosomal RNA clones (must be <1%), etc.) and to provide a sequence database representing the predominant clones in each library. The clones were sequenced at the University of Missouri-Columbia DNA Core Facility. Bioinformatics work was performed by GK Springer's bioinformatics group (WG Spollen, JE Ries, A Guillen, AA Khambati, RV Patel, CM Topinka, SB Bhuiyan) in Computer Science and Health Management and Informatics Departments at the University of Missouri-Columbia. Clone Requests: Requests for clones should be made to the Director of the University of Missouri DNA Core facility at: porcine@net.missouri.edu. Citations: 1. Banaldo MF, Lennon G, Soares MB. Normalization and Subtraction: Two approaches to facilitate gene discovery. Genome Res, 1996; 6:791-806. 2. Jiang H, Bivens NJ, Ries JE, Whitworth KM, Green JA, Forrester LJ, Springer GK, Didion BA, Mathialagan N, Prather RS, Lucy MC (2001) Constructing cDNA libraries with fewer clones that contain long poly(dA) tails. Biotechniques 31:38-42. 3. Soares MB, MF Banaldo, P Jelene, L Su, L Lawton, A Estrandiadis. 1994. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci, 91:9228-9232. TAG_TISSUE=Perinatal ovary TAG_SEQ=Not found"

ORIGIN

Query Match 73.6%; Score 36.8; DB 7; Length 425;
Best Local Similarity 85.4%; Pred. No. 0.001;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCGTACTGCCGCTTTCCAGTCGGGAACCTGTGCG 50
|||||
Db 219 TTAATTGCTAGCGCTACTGCCGCTTTCCAGTCGGGAACCTGTGCG 266
|||||

RESULT 9
BE403103
LOCUS

451 bp mRNA linear EST 21-JUL-2000

DEFINITION GBX002.C11F990602#08 ITEC GBX Wheat Root Library Triticum aestivum
cDNA clone GBX002.C11, mRNA sequence.

ACCESSION BE403103
VERSION BE403103.1 GI:9362483
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 451)
AUTHORS Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.
TITLE International Triticeae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triticeae
JOURNAL Unpublished (2000)
COMMENT Contact: Jacquemin JM
Centre de Recherches Agronomiques, Departement de Biotechnologie
234 chaussée de Charleroi, 5030 Gembloux BELGIUM
Tel: 32 81 61 29 35
Fax: 32 81 61 04 59
Email: jacquemin@cragx.fgov.be
International Triticeae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers
1. .451
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Odeon"
/db_xref="taxon:4565"
/clone="GBX002.C11"
/tissue_type="root"
/dev_stage="seedling, unstressed"
/clone_lib="ITEC GBX Wheat Root Library"
/note="Vector: pUC18; 0.3-2.0 Kbp average insert size."

ORIGIN
Query Match 73.6%; Score 36.8; DB 2; Length 451;
Best Local Similarity 85.4%; Pred. No. 0.001;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCATGCTCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
|||||
Db 94 TTAATTTGTTGCTCACTCCCGCTTTCCAGTCGGGAAACCTGTGCG 141
|||||

RESULT 10
CNS07H85 455 bp DNA linear GSS 02-OCT-2001
LOCUS Anopheles gambiae GSS T7 end of clone 25H23 of library NotreDamel
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.
AL610727
AL610727.1 GI:15916912
GSS.
ORGANISM Anopheles gambiae (African malaria mosquito)
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

REFERENCE 1 (bases 1 to 455)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (01-OCT-2001) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
2 Web : www.genoscope.cns.fr
3 (bases 1 to 455)
AUTHORS Roth,C.W., Brey,P.T., Ke,Z. and Collins,F.H.
TITLE Direct Submission

JOURNAL Submitted (01-OCT-2001) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.

FEATURES
source
1. .455
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="25H23"
/clone_lib="NotreDamel"
/note="end : T7"

ORIGIN
Query Match 73.6%; Score 36.8; DB 9; Length 455;
Best Local Similarity 85.4%; Pred. No. 0.001;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCATGCTCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
|||||
Db 154 TTAATTCGTTGCTCACTCCCGCTTTCCAGTCGGGAAACCTGTGCG 201
|||||

RESULT 11
LOCUS CO048749 591 bp mRNA linear EST 14-JUN-2004
DEFINITION tk66a07.b7 Arabidopsis RT-PCR Products (CSHL) Arabidopsis thaliana cDNA clone tk66a07, mRNA sequence.
ACCESSION CO048749
VERSION CO048749.1 GI:48688371
KEYWORDS EST.
ORGANISM Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 591)
AUTHORS Katari,M.S., Balija,V. and McCombie,W.R.
JOURNAL Arabidopsis RT-PCR Products (CSHL)
COMMENT Contact: Manpreet S. Katari
Woodbury Genome Center
Cold Spring Harbor Laboratory
500 Sunnyside Blvd., Plainview, NY 11797, USA
Tel: 516 422 4086
Fax: 516 422 4109
Email: mcombie@cshl.org
Matches At4g04920
Plate: tk66 row: a column: 07
High quality sequence stop: 591.
Location/Qualifiers
1. .591
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="tk66a07"
/tissue_type="Whole Plant"
/clone_lib="Arabidopsis RT-PCR Products (CSHL)"
/note="DNA was extracted from Arabidopsis thaliana whole plant tissue, provided by members of Rob Martienssen's lab, using TRIzol. Primers were designed in Hypothetical genes and un-annotated regions in Arabidopsis that are conserved in Brassica oleracea located in the short arm of chromosome 4. PCR products were either cloned into PCR TOPO 2.1 vector (Invitrogen) and then sequenced using -21 M13 forward and reverse universal primers or treated with Exonuclease I and Shrimp Alkaline phosphatase and sequenced using the specific primers."

FEATURES
source
1. .591
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="tk66a07"
/tissue_type="Whole Plant"
/clone_lib="Arabidopsis RT-PCR Products (CSHL)"
/note="DNA was extracted from Arabidopsis thaliana whole plant tissue, provided by members of Rob Martienssen's lab, using TRIzol. Primers were designed in Hypothetical genes and un-annotated regions in Arabidopsis that are conserved in Brassica oleracea located in the short arm of chromosome 4. PCR products were either cloned into PCR TOPO 2.1 vector (Invitrogen) and then sequenced using -21 M13 forward and reverse universal primers or treated with Exonuclease I and Shrimp Alkaline phosphatase and sequenced using the specific primers."

ORIGIN


```

Query Match      73.6%; Score 36.8; DB 7; Length 591;
Best Local Similarity 85.4%; Pred. No. 0.0011;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TTAAGCTTATACGATGACTGCCCGCTTTCCAGTCGGGAACCTGTCG 50
|||||
Db 521 TTAATTGCGTAGCGTCACTGCCCGCTTTCCAGTCGGGAACCTGTCG 568
|||||

RESULT 12
AG068743
LOCUS
DEFINITION
Pan troglodytes DNA, clone: PTB-059B24.F, genomic survey sequence.
AG068743
ACCESSION
AG068743.1 GI:16620545
VERSION
GSS.
KEYWORDS
Pan troglodytes (chimpanzee)
SOURCE
Pan troglodytes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1
REFERENCE
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
TITLE
JOURNAL
Unpublished
REFERENCE
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
COMMENT
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..655
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-041G18.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match      73.6%; Score 36.8; DB 9; Length 655;
Best Local Similarity 85.4%; Pred. No. 0.0011;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TTAAGCTTATACGATGACTGCCCGCTTTCCAGTCGGGAACCTGTCG 50
|||||
Db 242 TTAATTGTTGGCTCACTGCCCGCTTTCCAGTCGGGAACCTGTCG 289
|||||

RESULT 14
AG125607
LOCUS
DEFINITION
Pan troglodytes DNA, clone: PTB-135O16.F, genomic survey sequence.
AG125607
ACCESSION
AG125607.1 GI:16654772
VERSION
GSS.
KEYWORDS
Pan troglodytes (chimpanzee)
SOURCE
Pan troglodytes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1
REFERENCE
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
BAC end sequences of Library PTB
TITLE
JOURNAL
Unpublished
REFERENCE
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
COMMENT
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..641
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-059B24.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match      73.6%; Score 36.8; DB 9; Length 641;
Best Local Similarity 85.4%; Pred. No. 0.0011;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TTAAGCTTATACGATGACTGCCCGCTTTCCAGTCGGGAACCTGTCG 50
|||||
Db 217 TTAATTGTTGGCTCACTGCCCGCTTTCCAGTCGGGAACCTGTCG 264
|||||

RESULT 13
AG055437
LOCUS
DEFINITION
Pan troglodytes DNA, clone: PTB-041G18.F, genomic survey sequence.
AG055437
ACCESSION
AG055437.1 GI:16592880
VERSION
GSS.
KEYWORDS
Pan troglodytes (chimpanzee)
SOURCE
Pan troglodytes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Sequencing: -21M13

LIBRARY

Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.

FEATURES

source

Location/Qualifiers

1..681
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-135016.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN

Query Match 73.6%; Score 36.8; DB 9; Length 681;
Best Local Similarity 85.4%; Pred. No. 0.0011;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 3 TTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
|||||
Db 255 TTAATTGTTTGGCGTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 302
|||||

Search completed: October 24, 2005, 21:53:31
Job time : 367.936 secs

RESULT 15

AG118925

LOCUS

AG118925 682 bp DNA linear GSS 03-NOV-2001
Pan troglodytes DNA, clone: PTB-127D11.F, genomic survey sequence.

DEFINITION

AG118925

ACCESSION

AG118925.1

VERSION

GI:16739444

KEYWORDS

GSS.

SOURCE

Pan troglodytes (chimpanzee)

ORGANISM

Pan troglodytes

REFERENCE

1

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 682)

Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

Totoki,Y., Watanabe,H. and Sakaki,Y.

Direct Submission

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:chimbesc@gsc.riken.go.jp, URL:http://bgp.gsc.riken.go.jp/,

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

LIBRARY

Sequencing: -21M13

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI.

FEATURES

source

Location/Qualifiers

1..682
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-127D11.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN

Query Match 73.6%; Score 36.8; DB 9; Length 682;
Best Local Similarity 85.4%; Pred. No. 0.0011;
Matches 41; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:39:18 ; Search time 54.6144 Seconds
(without alignments)
5419.578 Million cell updates/sec

Title: US-09-896-888A-10
Perfect score: 50
Sequence: 1 acttaagctttatagcgtga.....tccagtcgggaacactgtcg 50

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	50	2	AAV62498 Plasmid p
2	36	72.0	16091	5	AA889978 DNA encod
3	36	72.0	20795	5	AA892596 DNA encod
4	35.8	71.6	745	3	AA87692 Human sec
5	35.8	71.6	791	11	ACN86091 Breast ca
6	35.8	71.6	869	11	ACN86091 Breast ca
7	35.6	71.2	327	4	AA85360 Sequence
8	35.6	71.2	586	5	ADL39252 Human ova
9	35.6	71.2	1637	5	AA877556 DNA encod
10	35.6	71.2	1695	5	AA877545 DNA encod
11	35.6	71.2	1942	5	AA869138 DNA encod
12	35.6	71.2	2346	5	AA877547 DNA encod
13	35.6	71.2	2424	5	AA887523 DNA encod
14	35.6	71.2	2710	5	AA877561 DNA encod
15	35.6	71.2	2710	10	ADE09733 Novel DNA
16	35.6	71.2	2757	5	AA869873 DNA encod
17	35.6	71.2	3859	12	ADE48102 Human che
18	35.6	71.2	5909	5	AA869432 DNA encod
19	35.6	71.2	10771	5	AA885906 DNA encod
20	35.6	71.2	20974	5	AA892595 DNA encod

21	35.2	70.4	60	4	AAF55402	Sequence
22	35.2	70.4	60	4	AAF55403	Sequence
23	35.2	70.4	114	10	ADL18713	Plasmid p
24	35.2	70.4	118	2	AAT01221	Oligonucle
25	35.2	70.4	151	2	AAV62171	HSV-2 str
26	35.2	70.4	173	5	ABV36253	Human pro
27	35.2	70.4	176	4	AAK85632	Human pro
28	35.2	70.4	179	5	ABV45247	Human pro
29	35.2	70.4	186	1	AAAN60799	Portion o
30	35.2	70.4	194	4	AAK85619	Human inn
31	35.2	70.4	195	4	AAK90169	Human dig
32	35.2	70.4	195	4	AAK89383	Human dig
33	35.2	70.4	195	4	AAK89976	Human dig
34	35.2	70.4	195	4	AAK79577	Human inn
35	35.2	70.4	195	4	AAK85615	Human inn
36	35.2	70.4	195	4	AAK85621	Human inn
37	35.2	70.4	195	4	AAK85626	Human inn
38	35.2	70.4	195	4	AAK85628	Human inn
39	35.2	70.4	195	4	AAK85633	Human inn
40	35.2	70.4	195	4	AAK85616	Human inn
41	35.2	70.4	195	4	AAK73946	Human inn
42	35.2	70.4	195	4	AAK69739	Human inn
43	35.2	70.4	195	4	AAK69742	Human inn
44	35.2	70.4	195	4	AAK85617	Human inn
45	35.2	70.4	195	4	AAK85618	Human inn

ALIGNMENTS

RESULT 1

AAV62498

ID AAV62498 standard; DNA; 50 BP.

XX AC AAV62498;

XX DT 19-JAN-1999 (first entry)

XX DE Plasmid p2Op2J-3 constructing primer 2.

XX KW Op ie2; promoter; shuttle vector; transformation; melanotransferrin;

XX KW immediate early baculovirus promoter; prokaryotic; transcription;

XX KW bleomycin/bleomycin-type antibiotic; insect cell; transposon;

XX KW ion transport peptide hormone; PCR primer; ss.

XX OS Synthetic.

XX FN WO9844141-A2.

XX PD 08-OCT-1998.

XX PF 26-MAR-1998; 98WO-CA000282.

XX PR 27-MAR-1997; 97US-0049946P.

XX PR 28-JAN-1998; 98CA-02221819.

XX PA (UYBR-) UNIV BRITISH COLUMBIA.

XX PI Grigliatti TA, Theilmann DA, Pfeifer TA, Hegedus DD;

XX DR WPI; 1998-557129/47.

XX PT Expression vectors for transforming insect cells from disparate lines -

XX PT useful to express heterologous DNA, e.g. to allow study of gene

XX PT expression and produce commercially important proteins.

XX PS Disclosure; Page 39; 121pp; English.

XX CC Primers AAV62497 and AAV62498 were used for the construction of the

XX CC plasmid p2Op2J-3. The invention provides a new shuttle vector for

XX CC transforming insect cells that comprises: (i) prokaryotic origin of

XX CC replication; (ii) insect promoter having homology to, and capable of

XX CC functioning as, an immediate early baculovirus promoter; (iii)

CC prokaryotic promoter sequence, and (iv) selectable marker capable of
 CC conferring resistance to a bleomycin/photomycin-type antibiotic under
 CC transcriptional control of (ii) and (iii), in insect and prokaryotic
 CC cells respectively. The vectors can be used to stably transform
 CC (especially insect) cells with heterologous DNA, is useful to allow study
 CC of gene expression and direct expression of heterologous gene products,
 CC such as commercially important proteins. They are especially useful to
 CC allow expression of melanotransferrins, ion transport peptide hormones or
 CC biologically active derivatives in insect cells. They enable
 CC transformation of insect cell lines from disparate species, allowing
 CC screening of lines for optimum post-translational modification of
 CC particular proteins. Shuttle vectors further comprising DNA transposable
 CC elements defining a transposon can be used to optimise heterologous
 CC protein expression and facilitate selection of desired transformants
 CC
 XX Sequence 50 BP; 11 A; 14 C; 12 G; 13 T; 0 U; 0 Other;
 SQ
 Query Match 100.0%; Score 50; DB 2; Length 50;
 Best Local Similarity 100.0%; Pred. No. 2.4e-11;
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ACTTAAGCTTATAGCATGACTGCCCGCTTCCAGTCGGGAAACCTGTGC 50
 Db 1 ACTTAAGCTTATAGCATGACTGCCCGCTTCCAGTCGGGAAACCTGTGC 50
 RESULT 2
 AAS89978/c
 ID AAS89978 standard; cDNA; 16091 BP.
 XX
 AC AAS89978;
 XX
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #25782.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 KW Homo sapiens.
 OS
 PN WO200175067-A2.
 FN
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG25791.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 25782; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS84197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: the sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 16091 BP; 4724 A; 3867 C; 3861 G; 3639 T; 0 U; 0 Other;
 Query Match 72.0%; Score 36; DB 5; Length 16091;
 Best Local Similarity 88.6%; Pred. No. 0.00011;
 Matches 39; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Qy 7 GCTTATAGCATGACTGCCCGCTTCCAGTCGGGAAACCTGTGC 50
 Db 8080 GTTATTGCAATCATTCGCCCGCTTCCAGTCGGGAAACCTGTGC 8037
 RESULT 3
 AAS92596
 ID AAS92596 standard; cDNA; 20795 BP.
 XX
 AC AAS92596;
 XX
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #28400.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 KW Homo sapiens.
 OS
 PN WO200175067-A2.
 FN
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG28409.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 28400; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostic, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 20795 BP; 4255 A; 5208 C; 5716 G; 5614 T; 0 U; 2 Other;

Query Match 72.0%; Score 36; DB 5; Length 20795;
Best Local Similarity 88.6%; Pred. No. 0.00011;
Matches 39; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 GCTTATAGCATGACTGCCCGCTTTCCAGTCGGGAACCTGTCG 50

Db 12363 GTTATTGCAATCATGTCCCGCTTTCCAGTCGGGAACCTGTCG 12406

RESULT 4

AAA87692/C

ID AAA87692 standard; cDNA; 745 BP.

XX AC AAA87692;

XX DT 04-DEC-2000 (first entry)

XX DE Human secreted protein gene 27 SEQ ID NO:37.

XX Human; secreted protein; immunosuppressive; immunostimulant; nootropic;
KW antiinflammatory; cardiac; vulnary; antiulcer; anticonvulsant;
KW antiparkinsonian; neuroprotective; antiviral; antibacterial; cytostatic;
KW antiparasitic; thrombolytic; anticoagulant; antiarteriosclerotic;
KW gene therapy; vaccine; chemotaxis-modulator; angiogenesis-modulator;
KW cancer; immune system disorder; hyperproliferative disorder; infection;
KW cardiovascular disorder; neurological disease; wound healing; ss.

OS Homo sapiens.

XX WO200043495-A2.

XX PD 27-JUL-2000.

XX PF 18-JAN-2000; 2000WO-US000903.

XX PR 19-JAN-1999; 99US-0116330P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM, Ebner R, Young PE, Ni J, Moore PA;

XX PI Komatsoulis G, Birse CE;

XX WPI; 2000-499225/44.

XX P-PSDB; AAB25691.

XX New isolated polynucleotide encoding a secreted protein useful for
PT preventing, treating or ameliorating a medical condition.

XX Claim 1; Page 394; 451pp; English.

XX The polynucleotide sequences given in AAA87666 to AAA87708 encodes the
CC human secreted proteins given in AAB25665 to AAB25755. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC immunostimulant; antiinflammatory; cardiac; vulnary; antiulcer;
CC nootropic; antiviral; anticonvulsant; antiparkinsonian; neuroprotective;
CC antibacterial; antiparasitic; thrombolytic; anticoagulant;
CC antiarteriosclerotic and cytostatic. The secreted proteins and their
CC polynucleotides can be used in gene therapy and as vaccines, chemotaxis-
CC modulators and angiogenesis- modulators. The human secreted proteins and

CC polynucleotides can be used for diagnosing (the susceptibility to) a
CC pathological condition by determining the presence or absence of a
CC mutation in the polynucleotide or determining the presence or amount of
CC expression of the protein. The polynucleotides and proteins can also be
CC used in the treatment and diagnosis of cancer, diseases of the immune
CC system, hyperproliferative disorders, cardiovascular disorders and
CC neurological disease. They can also be used to promote wound healing and
CC to fight infection. AAA87657 to AAA87665 and AAB25664 represent sequences
XX used in the exemplification of the present invention

SQ Sequence 745 BP; 191 A; 214 C; 188 G; 142 T; 0 U; 10 Other;

Query Match 71.6%; Score 35.8; DB 3; Length 745;
Best Local Similarity 83.3%; Pred. No. 5.9e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 3 TTAAGCTTATAGCATGACTGCCCGCTTTCCAGTCGGGAACCTGTCG 50

Db 168 TTAATNGCGTTGCGCTACTGCCCGCTTTCCAGTCGGGAACCTGTCG 121

RESULT 5

ACN86091

ID ACN86091 standard; DNA; 791 BP.

XX AC ACN86091;

XX DT 02-DEC-2004 (first entry)

XX DE Breast cancer related marker, seq id 7241.

XX Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.

OS Homo sapiens.

XX US2003099974-A1.

XX PD 29-MAY-2003.

XX PF 18-JUL-2002; 2002US-00198846.

XX PR 18-JUL-2001; 2001US-0306220P.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX WPI; 2003-787014/74.

XX Novel isolated polypeptide associated with breast cancer, useful for
PT detecting presence of polypeptide in sample, as a marker for breast
PT cancer.

XX Disclosure; SEQ ID NO 7241; 36pp; English.

XX The invention relates to an isolated polypeptide (I) associated with
CC breast cancer which is encoded by a nucleic acid molecule comprising a
CC nucleotide sequence (S1). Further disclosed is an antibody that binds to
CC the polypeptide of the invention. The activity of the polypeptide of the
CC invention may be described as cytostatic. The antibody is useful for
CC detecting the presence of (I) in a sample. Nucleic acid molecules of the
CC invention are useful in the detection of breast tumours. (I) is useful as
CC a marker for breast cancer and in breast cancer therapy. Sequences given
CC in records ACN78851-ACN92934 represent nucleic acid markers associated
CC with breast cancer. Note: The sequence listing does not form part of the
CC specification but may be obtained in electronic format from the USPTO web
CC site at seqdata.uspto.gov/sequence.html?DocID=20030099974

SQ Sequence 791 BP; 188 A; 198 C; 211 G; 182 T; 0 U; 12 Other;

Query Match 71.6%; Score 35.8; DB 11; Length 791;
Best Local Similarity 83.3%; Pred. No. 6e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

PR 21-MAR-2000; 2000US-0191031P.
 PR 25-MAY-2000; 2000US-0207124P.
 PR 15-JUN-2000; 2000US-0211940P.
 PR 07-JUL-2000; 2000US-0216820P.
 PR 25-JUL-2000; 2000US-0220661P.
 PR 21-DEC-2000; 2000US-0257672P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Lee J, Lillie J;
 XX
 DR WPI; 2001-611502/70.
 XX
 PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
 PT cancer cells as compared to their normal non-cancerous ovarian cells are
 PT used to characterize stage, grade, histological type of ovarian cancer.
 XX
 PS Disclosure; SEQ ID NO 13142; 106pp; English.
 XX
 CC The invention relates to nucleic acid markers which are overexpressed in
 CC ovarian cancer cells as compared to their expression in normal (i.e. non-
 CC cancerous) ovarian cells. The invention also relates to polypeptides
 CC encoded by the markers, antibodies that selectively bind to the
 CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
 CC of developing ovarian cancer involving inhibiting expression of a gene
 CC corresponding to a marker of the invention and a method of treating a
 CC patient afflicted with ovarian cancer comprising providing to cells of
 CC the patient an antisense oligonucleotide complementary to a marker of the
 CC invention. The markers are useful for assessing if a patient is afflicted
 CC with ovarian cancer, which involves comparing the level of expression of
 CC a marker in a patient sample and a normal level of expression of the
 CC marker in a control non-ovarian cancer sample. A difference between the
 CC expression levels indicates ovarian cancer. The level of expression of a
 CC marker corresponds to a secreted protein or to a transcribed
 CC polynucleotide or its portion. The level of expression of the marker is
 CC assessed by detecting the presence in the sample, a protein or protein
 CC fragment corresponding to the marker. The presence of protein or protein
 CC fragment is detected using an antibody that specifically binds with the
 CC protein or protein fragment. Alternatively, the level of expression of
 CC the marker is assessed by detecting the presence of a transcribed
 CC polynucleotide which anneals with the marker or anneals with a portion of
 CC the polynucleotide comprising the marker, under stringent conditions. The
 CC marker is also used for monitoring the progression of ovarian cancer in a
 CC patient which involves detecting expression of the marker in a patient
 CC sample at a first point in time, repeating the method at a subsequent
 CC time and comparing the level of expression. The method is carried out
 CC using an ovarian tissue sample. A composition comprising a marker,
 CC polypeptide or antibody of the invention is used to treat ovarian cancer.
 CC This sequence represents a human ovarian cancer DNA marker of the
 CC invention.
 XX
 SQ Sequence 586 BP; 168 A; 131 C; 141 G; 146 T; 0 U; 0 Other;
 XX
 Query Match 71.2%; Score 35.6; DB 5; Length 586;
 Best Local Similarity 82.0%; Pred. No. 6.7e-05;
 Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 ACTTAAGCTTATAGCAGTACGTCGCCGCTTTCCAGTCGCGAAACCTGCG 50
 516 ATTTAATGCTGTGCTGCTACGTCGCCGCTTTCCAGTCGCGAAACCTGCG 565
 DB
 RESULT 9
 AAS77556/c
 ID AAS77556 standard; cDNA; 1637 BP.
 XX
 AC AAS77556;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #13360.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR
 PR 23-AUG-2000; 2000US-00649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR
 DR P-PSDB; ABG13369.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 13360; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1637 BP; 447 A; 437 C; 395 G; 357 T; 0 U; 1 Other;
 XX
 Query Match 71.2%; Score 35.6; DB 5; Length 1637;
 Best Local Similarity 82.0%; Pred. No. 8.7e-05;
 Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 1 ACTTAAGCTTATAGCAGTACGTCGCCGCTTTCCAGTCGCGAAACCTGCG 50
 359 AATTAATGCTGTGCTGCTACGTCGCCGCTTTCCAGTCGCGAAACCTGCG 310
 DB
 RESULT 10
 AAS77545/c
 ID AAS77545 standard; cDNA; 1695 BP.
 XX
 AC AAS77545;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #13349.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX

OS Homo sapiens.
 XX WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG13358.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity.
 XX Claim 1; SEQ ID NO 13349; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 XX sequences. (I) is useful as hybridisation probes, polymerase chain
 XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 XX and in recombinant production of (II). The polynucleotides are also used
 XX in diagnostics as expressed sequence tags for identifying expressed
 XX genes. (I) is useful in gene therapy techniques to restore normal
 XX activity of (II) or to treat disease states involving (II). (II) is
 XX useful for generating antibodies against it, detecting or quantitating a
 XX polypeptide in tissue, as molecular weight markers and as a food
 XX supplement. (II) and its binding partners are useful in medical imaging
 XX of sites expressing (II). (I) and (II) are useful for treating disorders
 XX involving aberrant protein expression or biological activity. The
 XX polypeptide and polynucleotide sequences have applications in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX and to produce other types of data and products dependent on DNA and
 XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 XX coding sequences of the invention. Note: The sequence data for this
 XX patent did not appear in the printed specification, but was obtained in
 XX electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Query Match 71.2%; Score 35.6; DB 5; Length 1695;
 XX Best Local Similarity 82.0%; Pred. No. 8.8e-05;
 XX Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
 Db 651 AATTAAATTCGCTTGGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 602
 RESULT 11
 AAS69138/c
 ID AAS69138 standard; cDNA; 1942 BP.
 XX AAS69138;
 XX 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #4942.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.

PN WO200175067-A2.
 XX 11-OCT-2001.
 XX 30-MAR-2001; 2001WO-US008631.
 XX 31-MAR-2000; 2000US-00540217.
 XX 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 XX P-PSDB; ABG04951.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity.
 XX Claim 1; SEQ ID NO 4942; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 XX sequences. (I) is useful as hybridisation probes, polymerase chain
 XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 XX and in recombinant production of (II). The polynucleotides are also used
 XX in diagnostics as expressed sequence tags for identifying expressed
 XX genes. (I) is useful in gene therapy techniques to restore normal
 XX activity of (II) or to treat disease states involving (II). (II) is
 XX useful for generating antibodies against it, detecting or quantitating a
 XX polypeptide in tissue, as molecular weight markers and as a food
 XX supplement. (II) and its binding partners are useful in medical imaging
 XX of sites expressing (II). (I) and (II) are useful for treating disorders
 XX involving aberrant protein expression or biological activity. The
 XX polypeptide and polynucleotide sequences have applications in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX and to produce other types of data and products dependent on DNA and
 XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 XX coding sequences of the invention. Note: The sequence data for this
 XX patent did not appear in the printed specification, but was obtained in
 XX electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Query Match 71.2%; Score 35.6; DB 5; Length 1942;
 XX Best Local Similarity 82.0%; Pred. No. 9.1e-05;
 XX Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Qy 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
 Db 1183 AATTAAATTCGCTTGGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 1134
 RESULT 12
 AAS77547
 ID AAS77547 standard; cDNA; 2346 BP.
 XX AAS77547;
 XX 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #13351.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 XX WO200175067-A2.


```
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR P-PSDB; ABG23336.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 13351; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2346 BP; 542 A; 590 C; 609 G; 605 T; 0 U; 0 Other;
Query Match 71.2%; Score 35.6; DB 5; Length 2346;
Best Local Similarity 82.0%; Pred. No. 9.6e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 ACTTAAGCTTATAGCGATGACTGCCGCTTCCAGTCGGGAAACCTGTGCG 50
Db 1672 AATTAATTGCGTTGCGCTCACTGCCGCTTCCAGTCGGGAAACCTGTGCG 1721
RESULT 13
AAS87523/c
ID AAS87523 standard; cDNA; 2424 BP.
XX
XX AAS87523;
AC
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #23327.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
```

```
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR P-PSDB; ABG23336.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 23327; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2424 BP; 528 A; 721 C; 738 G; 437 T; 0 U; 0 Other;
Query Match 71.2%; Score 35.6; DB 5; Length 2424;
Best Local Similarity 82.0%; Pred. No. 9.7e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 ACTTAAGCTTATAGCGATGACTGCCGCTTCCAGTCGGGAAACCTGTGCG 50
Db 1365 AATTAATTGCGTTGCGCTCACTGCCGCTTCCAGTCGGGAAACCTGTGCG 1316
RESULT 14
AAS77561/c
ID AAS77561 standard; cDNA; 2710 BP.
XX
XX AAS77561;
AC
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #13365.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
```

PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG13374.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 1; SEQ ID NO 13365; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence novel data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 2710 BP; 615 A; 724 C; 690 G; 681 T; 0 U; 0 Other;

Query Match 71.2%; Score 35.6; DB 5; Length 2710;
Best Local Similarity 82.0%; Pred. No. 9,9e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 1080 AATTAATTGCGTTGGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 1031

RESULT 15
ADE09733/c
ID ADE09733 standard; DNA; 2710 BP.

XX ADE09733;

XX 29-JAN-2004 (first entry)

XX Novel DNA-related contig nucleotide sequence #455.

XX novel gene; novel protein; tissue marker; molecular weight marker;
XX chromosome marker; genetic disorder; contig; ds.

XX Unidentified.

XX WO2003054152-A2.

XX 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

XX 11-DEC-2001; 2001US-0339453P.

PR 14-MAR-2002; 2002US-0365091P.
PR 14-MAR-2002; 2002US-0365384P.
PR 12-APR-2002; 2002US-0372381P.
PR 12-APR-2002; 2002US-0372615P.
PR 22-APR-2002; 2002US-00128558.
PR 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.

XX Disclosure; SEQ ID NO 2277; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present DNA sequence was used in the
CC exemplification of the invention.

XX SQ Sequence 2710 BP; 615 A; 724 C; 690 G; 681 T; 0 U; 0 Other;

Query Match 71.2%; Score 35.6; DB 10; Length 2710;
Best Local Similarity 82.0%; Pred. No. 9.9e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 1080 AATTAATTGCGTTGGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 1031

Search completed: October 24, 2005, 18:58:10

Job time : 59.6144 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 20:18:24 ; Search time 102.023 Seconds
(without alignments)
4044.488 Million cell updates/sec

Title: US-09-896-888A-10
Perfect score: 50
Sequence: 1 acttaagttatagcgtga.....tccagtcgggaacctgtcg 50

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9772363 seqs, 4126298632 residues

Total number of hits satisfying chosen parameters: 19544726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
- 23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	50	US-09-896-888A-10	Sequence 10, Appl
c 2	36	72.0	16091	24	US-10-450-763-25782
3	36	72.0	20795	24	US-10-450-763-28400
c 4	35.8	71.6	745	11	US-09-985-153-37
5	35.8	71.6	791	15	US-10-198-846-7241

c 6	35.8	71.6	869	15	US-10-198-846-4171	Sequence 4171, Ap
c 7	35.6	71.2	415	20	US-10-437-963-80752	Sequence 80752, A
c 8	35.6	71.2	586	10	US-09-814-133-13142	Sequence 13142, A
c 9	35.6	71.2	865	14	US-10-027-632-2086	Sequence 2086, Ap
c 10	35.6	71.2	865	14	US-10-027-632-2087	Sequence 2087, Ap
c 11	35.6	71.2	865	18	US-10-027-632-2086	Sequence 2086, Ap
c 12	35.6	71.2	865	18	US-10-027-632-2087	Sequence 2087, Ap
c 13	35.6	71.2	1637	24	US-10-450-763-13360	Sequence 13360, A
c 14	35.6	71.2	1695	24	US-10-450-763-13349	Sequence 13349, A
c 15	35.6	71.2	1942	24	US-10-450-763-4942	Sequence 4942, Ap
c 16	35.6	71.2	2307	20	US-10-437-963-88806	Sequence 88806, A
c 17	35.6	71.2	2346	24	US-10-450-763-13351	Sequence 13351, A
c 18	35.6	71.2	2424	24	US-10-450-763-23327	Sequence 23327, A
c 19	35.6	71.2	2710	24	US-10-450-763-13365	Sequence 13365, A
c 20	35.6	71.2	2721	24	US-10-437-963-39259	Sequence 39259, A
c 21	35.6	71.2	2757	24	US-10-450-763-5677	Sequence 5677, Ap
c 22	35.6	71.2	5909	24	US-10-450-763-5236	Sequence 5236, Ap
c 23	35.6	71.2	10771	24	US-10-450-763-21710	Sequence 21710, A
c 24	35.6	71.2	20974	21	US-10-450-763-28399	Sequence 28399, A
c 25	35.2	70.4	173	21	US-10-357-930-36271	Sequence 36271, A
c 26	35.2	70.4	179	21	US-10-357-930-45266	Sequence 45266, A
c 27	35.2	70.4	195	9	US-09-764-846-347	Sequence 347, App
c 28	35.2	70.4	195	10	US-09-764-872-710	Sequence 710, App
c 29	35.2	70.4	195	10	US-09-764-891-7492	Sequence 7492, Ap
c 30	35.2	70.4	195	10	US-09-764-891-9818	Sequence 9818, Ap
c 31	35.2	70.4	195	15	US-10-091-483-347	Sequence 347, App
c 32	35.2	70.4	195	16	US-10-205-438-959	Sequence 959, App
c 33	35.2	70.4	201	17	US-10-016-986-41	Sequence 41, Appl
c 34	35.2	70.4	201	18	US-10-273-973-114	Sequence 114, Appl
c 35	35.2	70.4	203	21	US-10-357-930-36213	Sequence 36213, A
c 36	35.2	70.4	205	21	US-10-333-872A-286	Sequence 286, App
c 37	35.2	70.4	213	9	US-09-969-617-1	Sequence 1, Appl
c 38	35.2	70.4	238	18	US-10-319-227A-8	Sequence 8, Appl
c 39	35.2	70.4	238	18	US-10-319-227A-44	Sequence 44, Appl
c 40	35.2	70.4	238	18	US-10-286-549A-8	Sequence 8, Appl
c 41	35.2	70.4	238	18	US-10-286-549A-44	Sequence 44, Appl
c 42	35.2	70.4	238	24	US-10-319-227A-8	Sequence 8, Appl
c 43	35.2	70.4	238	24	US-10-319-227A-44	Sequence 44, Appl
c 44	35.2	70.4	244	21	US-10-357-930-45545	Sequence 45545, A
c 45	35.2	70.4	253	21	US-10-357-930-38331	Sequence 38331, A

ALIGNMENTS

RESULT 1
US-09-896-888A-10
; Sequence 10, Application US/09896888A
; Patent No. US20020116723A1
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Insect Expression Vectors
; FILE REFERENCE: 80021-44
; CURRENT APPLICATION NUMBER: US/09/896,888A
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: US/09/048,911
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/049,946
; PRIOR FILING DATE: 1997-03-27
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer
US-09-896-888A-10

Query Match 100.0%; Score 50; DB 9; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.4e-11;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50

RESULT 2

US-10-450-763-25782/c
; Sequence 25782, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 25782
; LENGTH: 16091
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (304)..(1047)
; OTHER INFORMATION: 99% homologous to Cloning vector pBACE3.6
; OTHER INFORMATION: levansucrase, accession number US0929, Smith-Waterman Score=1333.
US-10-450-763-25782

Query Match 72.0%; Score 36; DB 24; Length 16091;
Best Local Similarity 88.6%; Pred. No. 7.8e-05;
Matches 39; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7 GCCTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 8080 GTTTATTGCAATCATTTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 8037

RESULT 3

US-10-450-763-28400
; Sequence 28400, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 28400
; LENGTH: 20795
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (961)..(899)
; OTHER INFORMATION: 100% homologous to Cloning vector pBACE3.6
; OTHER INFORMATION: levansucrase, accession number US0929, Smith-Waterman Score=110.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(20795)

; OTHER INFORMATION: n = a,t,c or g
US-10-450-763-28400

Query Match 72.0%; Score 36; DB 24; Length 20795;
Best Local Similarity 88.6%; Pred. No. 8.2e-05;
Matches 39; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 7 GCCTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 12363 GTTTATTGCAATCATTTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 12406

RESULT 4

US-09-985-153-37/c
; Sequence 37, Application US/09985153
; Publication No. US20040181047A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 33 Human Secreted Proteins
; FILE REFERENCE: P2036P1
; CURRENT APPLICATION NUMBER: US/09/985,153
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: 09/618,150
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: PCT/US00/00903
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: 60/116,330
; PRIOR FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (3)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (48)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (93)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (113)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (163)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (727)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (739)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (745)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-985-153-37

Query Match 71.6%; Score 35.8; DB 11; Length 745;
Best Local Similarity 83.3%; Pred. No. 5.4e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 168 TTAATNGCGTTGCGCTACTACCCCGCTTTCCAGTCGGGAAACCTGTGCG 121

```
RESULT 5
US-10-198-846-7241
; Sequence 7241, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7241
; LENGTH: 791
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 282, 300, 554, 592, 650, 707, 720, 728, 755, 767, 775, 791
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-7241

Query Match 71.6%; Score 35.8; DB 15; Length 791;
Best Local Similarity 83.3%; Pred. No. 5.4e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3 TTAAGCTTATAGCGATGACTGCCGCTTTCCAGTCGGGAAACCTGTGCG 50
|||||
Db 544 TTAATGCGTNGCGTCACTGCCGCTTTCCAGTCGGGAAACCTGTGCG 591
|||||

RESULT 6
US-10-198-846-4171
; Sequence 4171, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4171
; LENGTH: 869
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2, 3, 4, 6, 9, 10, 11, 531, 585, 667, 691, 696, 732, 742,
; LOCATION: 763, 764, 780, 816, 841, 856
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-4171

Query Match 71.6%; Score 35.8; DB 15; Length 869;
Best Local Similarity 83.3%; Pred. No. 5.5e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3 TTAAGCTTATAGCGATGACTGCCGCTTTCCAGTCGGGAAACCTGTGCG 50
|||||
Db 544 TTAATGCGTNGCGTCACTGCCGCTTTCCAGTCGGGAAACCTGTGCG 591
|||||

RESULT 7
US-10-437-963-80752/c
; Sequence 80752, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 80752
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_80346C.1
US-10-437-963-80752

Query Match 71.2%; Score 35.6; DB 20; Length 415;
Best Local Similarity 82.0%; Pred. No. 5.9e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ACTTAAGCTTATAGCGATGACTGCCGCTTTCCAGTCGGGAAACCTGTGCG 50
|||||
Db 148 AATCATGTCATAGCTGTTTCTGCCGCTTTCCAGTCGGGAAACCTGTGCG 99
|||||

RESULT 8
US-09-814-353-13142
; Sequence 13142, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13142
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-13142
```


OTHER INFORMATION: 100% homologous to Homo sapiens endoglycan, accession number AF219137, Smith-Waterman Score=1070.

FEATURE:

NAME/KEY: misc feature
LOCATION: (1)...(1637)
OTHER INFORMATION: n = a,t,c or g
US-10-450-763-13360

Query Match 71.2%; Score 35.6; DB 18; Length 865;
Best Local Similarity 82.0%; Pred. No. 6.7e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

US-10-450-763-13360/c

Sequence 13360, Application US/10027632

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-07-12

PRIOR FILING DATE: 2000-04-20

PRIOR FILING DATE: 2000-03-29

PRIOR FILING DATE: 2000-02-24

PRIOR FILING DATE: 1999-11-23

PRIOR FILING DATE: 1999-09-28

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2087

LENGTH: 865

TYPE: DNA

ORGANISM: Human

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)...(865)

OTHER INFORMATION: n = A,T,C or G

US-10-027-632-2087

Query Match 71.2%; Score 35.6; DB 18; Length 865;
Best Local Similarity 82.0%; Pred. No. 6.7e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

US-10-450-763-13360/c

Sequence 13360, Application US/10450763

Publication No. US20050196754A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc

TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 790CIP3/US

CURRENT FILING DATE: 2003-06-11

PRIOR FILING DATE: 2001-03-30

PRIOR FILING DATE: 2000-03-31

PRIOR FILING DATE: 2000-03-31

PRIOR FILING DATE: 2000-08-23

NUMBER OF SEQ ID NOS: 60736

SOFTWARE: Custom

SEQ ID NO 13360

LENGTH: 1637

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SIMILAR

LOCATION: (241)..(849)

US-10-027-632-2087

Query Match 71.2%; Score 35.6; DB 24; Length 1695;
Best Local Similarity 82.0%; Pred. No. 7.6e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

US-10-450-763-13349/c

Sequence 13349, Application US/10450763

Publication No. US20050196754A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc

TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 790CIP3/US

CURRENT FILING DATE: 2003-06-11

PRIOR FILING DATE: 2001-03-30

PRIOR FILING DATE: 2000-03-31

PRIOR FILING DATE: 2000-08-23

NUMBER OF SEQ ID NOS: 60736

SOFTWARE: Custom

SEQ ID NO 13360

LENGTH: 1637

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SIMILAR

LOCATION: (241)..(849)

US-10-027-632-2087

; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 4942
; LENGTH: 1942
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (1514)..(1942)
; OTHER INFORMATION: 100% homologous to Cloning vector pSacBII SacB, accession
; OTHER INFORMATION: number U09128, Smith-Waterman Score=768.
US-10-450-763-4942

Query Match 71.2%; Score 35.6; DB 24; Length 1942;
Best Local Similarity 82.0%; Pred. No. 7.8e-05;
Matches 41; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Oy 1 ACTTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 50
Db 1183 AATTAATTCGCTTGGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCG 1134

Search completed: October 25, 2005, 06:13:42
Job time : 106.689 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2005, 18:41:53 ; Search time 17.0038 Seconds
(without alignments)
4811.505 Million cell updates/sec

Title: US-09-896-888A-10

Perfect score: 50

Sequence: 1 acttaagttatagcgtga.....tccagtcgggaacctgtcg 50

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/prodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/prodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/prodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/prodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
C 1	35.2	70.4	114	2	US-08-778-217-1 Sequence 1, Appli
C 2	35.2	70.4	114	2	US-08-821-948-1 Sequence 1, Appli
C 3	35.2	70.4	114	4	US-09-397-955C-1 Sequence 1, Appli
C 4	35.2	70.4	118	5	PCT-US95-04092-1 Sequence 1, Appli
C 5	35.2	70.4	201	1	US-08-276-852-41 Sequence 41, Appl
C 6	35.2	70.4	201	1	US-08-133-011-114 Sequence 114, App
C 7	35.2	70.4	201	1	US-08-322-730A-114 Sequence 114, App
C 8	35.2	70.4	201	1	US-08-387-874-87 Sequence 87, Appl
C 9	35.2	70.4	201	1	US-08-899-575-41 Sequence 41, Appl
C 10	35.2	70.4	201	1	US-08-899-575-41 Sequence 41, Appl
C 11	35.2	70.4	201	2	US-08-383-619-114 Sequence 114, App
C 12	35.2	70.4	201	3	US-08-307-739-114 Sequence 114, App
C 13	35.2	70.4	201	3	US-09-729-597-114 Sequence 114, App
C 14	35.2	70.4	201	5	PCT-US93-08364-87 Sequence 87, Appl
C 15	35.2	70.4	201	5	PCT-US95-08743-41 Sequence 41, Appl
C 16	35.2	70.4	221	1	US-08-531-132-1 Sequence 1, Appli
C 17	35.2	70.4	304	4	US-08-120-324-15 Sequence 15, Appl
C 18	35.2	70.4	310	4	US-10-165-857-1 Sequence 1, Appli
C 19	35.2	70.4	310	4	US-10-165-856A-1 Sequence 1, Appli
C 20	35.2	70.4	352	4	US-10-165-857-2 Sequence 2, Appli
C 21	35.2	70.4	352	4	US-10-165-856A-2 Sequence 2, Appli
C 22	35.2	70.4	357	3	US-09-525-046-3 Sequence 3, Appli
C 23	35.2	70.4	360	4	US-08-182-173A-1 Sequence 1, Appli
C 24	35.2	70.4	450	4	US-09-486-336A-2 Sequence 2, Appli
C 25	35.2	70.4	504	2	US-08-768-550-12 Sequence 12, Appl
C 26	35.2	70.4	505	2	US-08-768-550-11 Sequence 11, Appl
C 27	35.2	70.4	506	2	US-08-768-550-10 Sequence 10, Appl

28	35.2	70.4	584	4	US-09-702-705-639	Sequence 639, App
29	35.2	70.4	584	4	US-09-736-457-639	Sequence 639, App
30	35.2	70.4	584	4	US-09-614-124B-639	Sequence 639, App
31	35.2	70.4	584	4	US-09-671-325-639	Sequence 639, App
32	35.2	70.4	584	4	US-09-589-184-639	Sequence 639, App
33	35.2	70.4	584	4	US-09-658-824-639	Sequence 639, App
34	35.2	70.4	585	4	US-09-334-818A-7	Sequence 7, Appli
35	35.2	70.4	591	4	US-09-334-818A-3	Sequence 10, Appl
36	35.2	70.4	597	4	US-09-334-818A-10	Sequence 16, Appl
37	35.2	70.4	597	4	US-09-334-818A-16	Sequence 16, Appl
38	35.2	70.4	598	4	US-09-334-818A-11	Sequence 11, Appl
39	35.2	70.4	599	4	US-09-334-818A-6	Sequence 6, Appli
40	35.2	70.4	601	4	US-09-334-818A-18	Sequence 18, Appl
41	35.2	70.4	601	4	US-09-334-818A-19	Sequence 19, Appl
42	35.2	70.4	602	4	US-09-334-818A-14	Sequence 14, Appl
43	35.2	70.4	602	4	US-09-334-818A-15	Sequence 15, Appl
44	35.2	70.4	602	4	US-09-334-818A-22	Sequence 22, Appl
45	35.2	70.4	604	4	US-09-334-818A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-778-217-1/c
; Sequence 1, Application US/08778217
; Patent No. 5935833
; GENERAL INFORMATION:
; APPLICANT: Kacian et al.
; TITLE OF INVENTION: Highly-Purified Recombinant
; TITLE OF INVENTION: Reverse Transcriptase
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 10210 Genetic Center Drive
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92121-4362
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Microsoft MS-DOS (Version 6.0)
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/778,217
; FILING DATE: January 9, 1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/221,804
; FILING DATE: April 1, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Christine A. Gritzmacher
; REGISTRATION NUMBER: 40,627
; REFERENCE/DOCKET NUMBER: MOL2A-A01F01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 410-8926
; TELEFAX: (619) 410-8928
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 114
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-778-217-1

Query Match 70.4%; Score 35.2; DB 2; Length 114;
Best Local Similarity 83.3%; Pred. No. 9.6e-06;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 3 TTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGAAACCTGTCG 50

Db 63 TTAATGCTGCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGCG 16
|||||
PRIOR FILING DATE: 1997-03-21
PRIOR APPLICATION NUMBER: 08/443,781
PRIOR FILING DATE: 1995-05-18
PRIOR APPLICATION NUMBER: 08/221,804
PRIOR FILING DATE: 1994-04-01
NUMBER OF SEQ ID NOS: 21
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 114
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
OTHER INFORMATION: Oligonucleotide used to construct plasmid pUC 18N
US-09-397-955C-1
Query Match 70.4%; Score 35.2; DB 2; Length 114;
Best Local Similarity 83.3%; Pred. No. 9.6e-06;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Oy 3 TTAAGCTTATAGCGATGACTCCCGCTTCCAGTCGGGAACCTGTGCG 50
|||||
Db 63 TTAATGCTGCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGCG 16
RESULT 4
PCT-US95-04092-1/c
Sequence 1, Application PC/TUS9504092
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HIGHLY-PURIFIED RECOMBINANT REVERSE
TITLE OF INVENTION: TRANSCRIPTASE
NUMBER OF SEQUENCES: 18
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04092
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 118 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-04092-1
Query Match 70.4%; Score 35.2; DB 5; Length 118;
Best Local Similarity 83.3%; Pred. No. 9.7e-06;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Oy 3 TTAAGCTTATAGCGATGACTCCCGCTTCCAGTCGGGAACCTGTGCG 50
|||||
Db 63 TTAATGCTGCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGCG 16
RESULT 5
US-08-276-852-41/c
Sequence 41, Application US/08276852
Patent No. 5652138
GENERAL INFORMATION:
APPLICANT: Burton, Dennis R
APPLICANT: Barbas, Carlos F
APPLICANT: Lerner, Richard A
TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
NUMBER OF SEQUENCES: 170
CORRESPONDENCE ADDRESS:

Db 63 TTAATGCTGCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGCG 16
|||||
PRIOR FILING DATE: 1997-03-21
PRIOR APPLICATION NUMBER: 08/443,781
PRIOR FILING DATE: 1995-05-18
PRIOR APPLICATION NUMBER: 08/221,804
PRIOR FILING DATE: 1994-04-01
NUMBER OF SEQ ID NOS: 21
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 114
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
OTHER INFORMATION: Oligonucleotide used to construct plasmid pUC 18N
US-09-397-955C-1
Query Match 70.4%; Score 35.2; DB 2; Length 114;
Best Local Similarity 83.3%; Pred. No. 9.6e-06;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Oy 3 TTAAGCTTATAGCGATGACTCCCGCTTCCAGTCGGGAACCTGTGCG 50
|||||
Db 63 TTAATGCTGCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGCG 16
RESULT 2
US-08-821-948-1/c
Sequence 1, Application US/08821948
Patent No. 5998195
GENERAL INFORMATION:
APPLICANT: Kacian et al.
TITLE OF INVENTION: Highly-Purified Recombinant
TITLE OF INVENTION: Reverse Transcriptase
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gen-Probe Incorporated
STREET: 10210 Genetic Center Drive
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92121-4362
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM compatible
OPERATING SYSTEM: Microsoft MS-DOS (Version 6.0)
SOFTWARE:
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/821,948
FILING DATE: March 22, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 08/443,781
FILING DATE: May 18, 1995
APPLICATION NUMBER: 08/221,804
FILING DATE: April 1, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Christine A. Gritzmacher
REGISTRATION NUMBER: 40,627
REFERENCE/DOCKET NUMBER: MOL2A (New Ref.: GP059-04.FW2)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 410-8926
TELEFAX: (619) 410-8928
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 114
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-821-948-1
Query Match 70.4%; Score 35.2; DB 2; Length 114;
Best Local Similarity 83.3%; Pred. No. 9.6e-06;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Oy 3 TTAAGCTTATAGCGATGACTCCCGCTTCCAGTCGGGAACCTGTGCG 50
|||||
Db 63 TTAATGCTGCGCTCACTCCCGCTTCCAGTCGGGAACCTGTGCG 16
RESULT 3
US-09-397-955C-1/c
Sequence 1, Application US/09397955C
Patent No. 6593120
GENERAL INFORMATION:
APPLICANT: RIGGS, Michael G.
APPLICANT: SORENSEN, Matthew
TITLE OF INVENTION: RECOMBINANT DNA ENCODING REVERSE TRANSCRIPTASE DERIVED FROM
TITLE OF INVENTION: MOLONEY MURINE LEUKEMIA VIRUS
FILE REFERENCE: GP059-05.CP1
CURRENT APPLICATION NUMBER: US/09/397,955C
CURRENT FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: 08/821,948

ADDRESSEE: The Scripps Research Institute, Office of
ADDRESS: Patent Counsel
STREET: 10666 No. 5652138th Torrey Pines Road, Suite 220,
STREET: Mail Drop TPC8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,852
FILING DATE: 18-JUL-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/178,302
FILING DATE: 30-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/954,148
FILING DATE: 30-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCR1452P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-276-852-41

Query Match 70.4%; Score 35.2; DB 1; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 3 TTAAGCTTATAGCGATGACTGCCCGCTTCCAGTCGGGAAACCTGTCG 50
Db 65 TTAATTGCGTTCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTCG 18

RESULT 6
US-08-133-011-114/c
Sequence 114 Application US/08133011
Patent No. 5658727
GENERAL INFORMATION:
APPLICANT: Kang, Angray
APPLICANT: Barbas, Carlos
APPLICANT: Lerner, Richard A.
TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING
NUMBER OF SEQUENCES: 161
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESS: Patent Counsel
STREET: 10666 No. 5658727th Torrey Pines Road, TPC-8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/133,011
FILING DATE: 08-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/683,602
FILING DATE: 10-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/826,623
FILING DATE: 27-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US 92/03091
FILING DATE: 10-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: SCRF 238.2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 201 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-133-011-114

Query Match 70.4%; Score 35.2; DB 1; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 3 TTAAGCTTATAGCGATGACTGCCCGCTTCCAGTCGGGAAACCTGTCG 50
Db 65 TTAATTGCGTTCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTCG 18

RESULT 7
US-08-322-730A-114/c
Sequence 114 Application US/08322730A
Patent No. 5759817
GENERAL INFORMATION:
APPLICANT: Barbas, Carlos
TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING
NUMBER OF SEQUENCES: 125
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Scripps Research Institute, Office of
ADDRESS: Patent Counsel
STREET: 10550 No. 5759817th Torrey Pines Road, Suite 220,
STREET: Mail Drop TPC8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/322,730A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/826,623
FILING DATE: 27-JAN-1992
APPLICATION NUMBER: US 07/683,602
FILING DATE: 10-APR-1991

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Bingham, Douglas A
; REGISTRATION NUMBER: 32,457
; REFERENCE/DOCKET NUMBER: SCR0707P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-784-2937
; TELEFAX: 619-784-9399
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-322-730A-114

Query Match          70.4%; Score 35.2; DB 1; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCATGACTCCCGCTTCCAGTCGGGAAACCTGTGC 50
    |||||
Db 65 TTAATTGGTTCGCTCACTCCCGCTTCCAGTCGGGAAACCTGTGC 18

RESULT 8
US-08-387-874-87/c
; Sequence 87, Application US/08387874
; Patent No. 5770356
; GENERAL INFORMATION:
; APPLICANT: Light, Paul L., II
; APPLICANT: Lerner, Richard A.
; TITLE OF INVENTION: PHAGEMIDS COEXPRESSING A SURFACE
; TITLE OF INVENTION: RECEPTOR AND A SURFACE HETEROLOGOUS PROTEIN
; NUMBER OF SEQUENCES: 97
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 5770356th Torrey Pines Road, Suite 220,
; STREET: Mail Drop TPC8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/387,874
; FILING DATE: 22-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08364
; FILING DATE: 03-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/941,369
; FILING DATE: 04-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI 303.1
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-387-874-87

Query Match          70.4%; Score 35.2; DB 1; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCATGACTCCCGCTTCCAGTCGGGAAACCTGTGC 50
    |||||
Db 65 TTAATTGGTTCGCTCACTCCCGCTTCCAGTCGGGAAACCTGTGC 18

RESULT 9
US-08-899-575-41/c
; Sequence 41, Application US/08899575
; Patent No. 5770440
; GENERAL INFORMATION:
; APPLICANT: Burton, Dennis R
; APPLICANT: Barbas, Carlos F
; APPLICANT: Lerner, Richard A
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
; NUMBER OF SEQUENCES: 170
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 5770440th Torrey Pines Road, Suite 220,
; STREET: Mail Drop TPC8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,575
; FILING DATE: 24-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/276,852
; FILING DATE: 18-JUL-1994
; APPLICATION NUMBER: US 08/178,302
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/954,148
; FILING DATE: 30-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCR1452P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-899-575-41

Query Match          70.4%; Score 35.2; DB 1; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
```

QY 3 TTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGC 50
|||||
Db 65 TTAATTGCGTTCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGC 18

RESULT 10

US-08-899-575-41/c
; Sequence 41, Application US/08899575
; Patent No. 5804440
; GENERAL INFORMATION:
; APPLICANT: Burton, Dennis R
; APPLICANT: Barbas, Carlos F
; APPLICANT: Lerner, Richard A
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
; NUMBER OF SEQUENCES: 170
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 580440th Torrey Pines Road, Suite 220,
; STREET: Mail Drop TPC8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,575
; FILING DATE: 24-JUL-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/276,852
; FILING DATE: 18-JUL-1994
; APPLICATION NUMBER: US 08/178,302
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/954,148
; FILING DATE: 30-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCR1452P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-899-575-41

Query Match 70.4%; Score 35.2; DB 1; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3 TTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGC 50
|||||
Db 65 TTAATTGCGTTCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGC 18

RESULT 11

US-08-383-619-114/c
; Sequence 114, Application US/08383619
; Patent No. 5955341

; GENERAL INFORMATION:
; APPLICANT: Kang, Angray
; APPLICANT: Barbas, Carlos
; APPLICANT: Lerner, Richard
; TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING
; TITLE OF INVENTION: PHAGEMIDS
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DOUGLAS A. BINGHAM
; STREET: 11300 Sorrento Valley Road, Suite 200
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/383,619
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/126,680
; FILING DATE:
; APPLICATION NUMBER: US/07/683,602
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bingham, Douglas A.
; REGISTRATION NUMBER: 32,457
; REFERENCE/DOCKET NUMBER: SCR0371P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-1555
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-383-619-114

Query Match 70.4%; Score 35.2; DB 2; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 3 TTAAGCTTATAGCGATGACTGCCCGCTTTCCAGTCGGGAAACCTGTGC 50
|||||
Db 65 TTAATTGCGTTCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGC 18

RESULT 12

US-08-907-739-114/c
; Sequence 114, Application US/08907739
; Patent No. 6235469
; GENERAL INFORMATION:
; APPLICANT: Kang, Angray
; APPLICANT: Barbas, Carlos
; APPLICANT: Lerner, Richard A.
; TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING
; TITLE OF INVENTION: PHAGEMIDS
; NUMBER OF SEQUENCES: 161
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 No. 6235469th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/907,739
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/133,011
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/826,623
; FILING DATE: 27-JAN-1992
; APPLICATION NUMBER: PCT/US 92/03091
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCRF 238.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-907-739-114

Query Match 70.4%; Score 35.2; DB 3; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCGATGACTCCCGCTTTCCAGTCGGGAAACCTGTGC 50
Db 65 TTAATTGCGTTGCGCTCACTCCCGCTTTCCAGTCGGGAAACCTGTGC 18

RESULT 13
US-09-729-597-114/c
; Sequence 114, Application US/09729597
; Patent No. 6468738
; GENERAL INFORMATION:
; APPLICANT: Kang, Angray
; Barbas, Carlos
; Lerner, Richard A.
; TITLE OF INVENTION: HETERODIMERIC RECEPTOR LIBRARIES USING
; PHAGEMIDS
; NUMBER OF SEQUENCES: 161
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; Patent Counsel
; STREET: 10666 No. 6468738th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/729,597
; FILING DATE: 04-Dec-2000
; CLASSIFICATION: 435
```

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/133,011
; FILING DATE: 1994-09-29
; APPLICATION NUMBER: US 07/826,623
; FILING DATE: 27-JAN-1992
; APPLICATION NUMBER: PCT/US 92/03091
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCRF 238.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 114:
; US-09-729-597-114

Query Match 70.4%; Score 35.2; DB 3; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCGATGACTCCCGCTTTCCAGTCGGGAAACCTGTGC 50
Db 65 TTAATTGCGTTGCGCTCACTCCCGCTTTCCAGTCGGGAAACCTGTGC 18

RESULT 14
PCT-US93-08364-87/c
; Sequence 87, Application PC/TUS9308364
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PHAGEMIDS COEXPRESSING A SURFACE
; RECEPTOR AND A SURFACE HETEROLOGOUS PROTEIN
; NUMBER OF SEQUENCES: 97
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08364
; FILING DATE: 03-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/941,369
; FILING DATE: 04-SEP-1992
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US93-08364-87

Query Match 70.4%; Score 35.2; DB 5; Length 201;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 TTAAGCTTATAGCGATGACTCCCGCTTTCCAGTCGGGAAACCTGTGC 50
Db 65 TTAATTGCGTTGCGCTCACTCCCGCTTTCCAGTCGGGAAACCTGTGC 18
```

RESULT 15
PCT-US95-08743-41/c
; Sequence 41, Application PC/TUS9508743
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES
; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS
; NUMBER OF SEQUENCES: 170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08743
; FILING DATE: 11-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/276,852
; FILING DATE: 18-JUL-1994
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-08743-41

Query Match 70.4%; Score 35.2; DB 5; Length 201;
Best Local Similarity 83.3%; Pred. NO. 1.1e-05;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 3 TTAAGCTTATACGATGACTGCCGCTTCCAGTCGGGAACCTGTCG 50
||| |
Db 65 TTAATGCGTTCGCTCACTGCCCGCTTCCAGTCGGGAACCTGTCG 18
||| |

Search completed: October 24, 2005, 21:57:51
Job time : 19.0038 secs

This Page Blank (uspto)